Internet Appendix for Speculation Sentiment

Shaun William Davies

IA.1 Internet Appendix Introduction

In this Internet Appendix, I provide support for the results in the main text. Section IA.2 extends the Brown, Davies and Ringgenberg (2021) model of the ETF mechanism to provide theoretical support for the empirical analysis. Section IA.3 addresses concerns about using higher frequency measures (e.g., daily or weekly). Section IA.4 details the small sample parametric bootstrap procedure used throughout the paper. Section IA.5 examines the relation between daily returns and a daily measure of SSI to show that the negative relation between SSI and contemporaneous returns holds even at higher frequencies. Section IA.6 shows analytically that maintaining a target level of portfolio leverage via leveraged ETFs requires a trading strategy that is contrarian. Section IA.6 also computes a measure of implied leverage rebalancing and shows that the return predictability results using SSI are robust to the inclusion of the measure. Section IA.7 provides a host of alternative specifications of SSI and considers an additional robustness test using changes in institutional ownership of leveraged ETFs. Section IA.8 studies the performance of portfolios that condition on realized values of the Speculation Sentiment Index. Section IA.9 provides the correlations between SSI and the control variables used throughout the paper.

IA.2 Model

In this section, I extend the Brown et al. (2021) model of the ETF mechanism to provide theoretical support for my empirical analysis. Specifically, the model extension shows that the Speculation Sentiment Index is a natural proxy for market-wide speculative demand shocks and the model extension also provides novel empirical predictions. There are four periods $t \in \{0, 1, 2, \textit{long term}\}$ in which two passively managed, leveraged ETFs are run by a risk neutral, competitive sponsor (e.g., ProShares). Both leveraged ETFs provide magnified exposure to a benchmark index χ (e.g., the S&P 500 index). One of the leveraged ETFs provides magnified *long* exposure to the benchmark index and the other leveraged ETF provides magnified *short* exposure. Furthermore, the long- and short-leverage are equal in magnitude, for example, the leveraged-long ETF provides $2\times$ exposure to the benchmark index and the leveraged-short ETF provides $-2\times$ exposure. I characterize the leveraged-long ETF with the subscript L and the leveraged-short ETF with the subscript S.

In each period t and for each ETF $j \in \{L, S\}$, there is a $q_{t,j}$-length measure of ETF shares traded in a centralized secondary market. The market value of each share is denoted $p_{t,j}$. The underlying assets backing each ETF are cash and a derivative contract (e.g., a total return swap) for which the reference entity is the index χ. The underlying derivative contracts for the leveraged-long and leveraged-short ETF are the same with the only difference being the position in the derivative contract; the leveraged-long ETF takes the long-end of the derivative contract (i.e, receives the total return on χ) and the leveraged-short ETF takes the short-end (i.e, pays the total return on χ). I make three assumptions regarding the derivative contract for tractability. First, I normalize the fixed rate/fee paid by the long-end to the short-end equal to zero. In other words, the net payoff of the derivative is the total return on χ. Second, I assume continuous settlement so that at any time t the fair value of the derivative position is equal to zero and the two parties' cash positions absorbs all gains and losses (this also alleviates concerns about roll yield). Third, I assume the derivative market is perfectly competitive and frictionless. As such, there are no arbitrage opportunities between the derivative contract and the underlying index χ.

Each unit of ETF j's underlying assets (cash plus the value of the derivative position) has a tradable value $\pi_{t,j}$ (i.e., net asset value or NAV). For simplicity, I assume that the number of units of the underlying asset are also $q_{t,j}$ so that the NAV per ETF share is $\pi_{t,j}$. Thus, in any period t, the *ETF premium* (or *ETF discount* when it is negative) for ETF j is the difference in the ETF's share price and the NAV per share,

$$\psi_{t,j} \equiv p_{t,j} - \pi_{t,j}. \tag{IA1}$$

$\psi_{t,j} \neq 0$ represents a *relative* mispricing (i.e., a violation of the law of one price) and an attractive opportunity for arbitrageurs.

In addition to the ETF shares trading in a secondary market, there also exists a primary market for each ETF j's shares. The primary market is constituted by $N \geq 1$ authorized participants (e.g., broker-dealers and market-makers) and the ETF sponsor.

Authorized participants are risk neutral and may be thought of as arbitrageurs. For simplicity, it is assumed that each authorized participant trades in both the leveraged-long and leveraged-short ETF markets. In response to $\psi_{t,j} > 0$, each authorized participant may short-sell ETF j's shares and hedge the short position with the underlying derivative security. In response to $\psi_{t,j} < 0$, each authorized participant may purchase ETF j's shares and hedge the long position with the underlying derivative security. After conducting the necessary long-short trade to exploit $\psi_{t,j} \neq 0$, an authorized participant closes the trade and captures the profits by creating or redeeming ETF j's shares. For example, if an authorized participant was short ETF shares, he may deliver cash in the amount of NAV in exchange for new ETF shares to cover the short position. Alternatively, if an authorized participant was long ETF shares, he may deliver those shares in exchange for cash in the amount of NAV. In either case, the authorized participant will simultaneously unwind their position in the derivative security used to hedge.[1]

Authorized participant trades only occur at $t = 2$, so no subscript t is needed on those variables. I denote each authorized participant i's arbitrage demand for ETF j as $\delta_{j,i}$. A positive value of $\delta_{L,i}$ implies shorting ETF L's shares, hedging with a long position in the derivative security, and subsequent share creations. A positive value of $\delta_{S,i}$ implies shorting ETF S's shares, hedging with a short position in the derivative security, and subsequent share creations. A negative value of $\delta_{L,i}$ implies buying ETF L's shares,

[1] In practice, leveraged ETFs adhere to a cash creation/redemption primary market. Conversely, most nonleveraged ETFs adhere to an in-kind creation/redemption primary market. Mathematically, in this model, it is equivalent to assume the transactions are in-kind. Furthermore, there is a nominal, fixed fee charged by the ETF sponsor for any creations or redemptions. I implicitly assume the fee is equal to zero for simplicity.

hedging with a short position in the derivative security, and subsequent share redemptions. A negative value of $\delta_{S,i}$ implies buying ETF S's shares, hedging with a long position in the derivative security, and subsequent share redemptions. The aggregate demand of authorized participants for ETF j is denoted as,

$$\Delta_j = \sum_{i=1}^{N} \delta_{j,i}. \tag{IA2}$$

The model's timing is as follows. Within period $t = 0$, initial prices for the leveraged ETFs' shares and the ETFs' NAVs are established, as are initial share quantities. Within period $t = 1$: i) demand shocks (both fundamental and nonfundamental) are realized, ii) investor demands shift for both ETFs' shares and the ETFs' underlying assets, giving rise to interim ETF premia. At $t = 2$: i) authorized participants trade against the premia (i.e., relative mispricing), generating price pressure on both ETFs' shares and their underlying assets, ii) authorized participants create or redeem shares in the ETFs to close their trades, and iii) the equilibrium prices for the ETFs' shares and NAVs are established. At $t =$ *long term*, prices return to their latent fundamental values. I elaborate below.

At $t = 0$, the initial ETF share price for ETF j is determined by a downward-sloped demand curve,

$$p_{0,j} = \beta_j - \eta_j q_{0,j}, \tag{IA3}$$

in which $\beta_j > 0$ is a constant and $\eta_j > 0$ proxies for investors' sensitivity to the measure of shares. A downward-sloped demand curve is micro-founded on investor risk aversion and lower values of η_j imply less price impact from changes in ETF shares outstanding. For simplicity, I assume that the ETFs' NAVs are initially equal,

$$\Omega \equiv \pi_{0,L} = \pi_{0,S}, \tag{IA4}$$

for some $\Omega > 0$. It is assumed that Ω is the true, latent fundamental value. Furthermore, I assume that no initial relative mispricing exist. That is,

$$q_{0,j} \equiv \frac{\beta_j - \Omega}{\eta_j}, \tag{IA5}$$

so that $p_{0,L} = p_{0,S} = \pi_{0,L} = \pi_{0,S} = \Omega$. In other words, all ETF shares and NAVs are efficiently priced at $t = 0$.

At $t = 1$, ETF j's share price is given by,

$$p_{1,j} = \beta_j - \eta_j q_{0,j} + (\omega + ssi)\,(1 - 2 \times \mathbb{1}_{j=S}), \tag{IA6}$$

in which $\mathbb{1}_{j=S}$ is an indicator function that equals 1 for the leveraged-short ETF (i.e., $j = S$) and equals 0 otherwise. ω is a fundamental shock related to the benchmark index (e.g., news about cash flows or discount rates) and ssi represents a nonfundamental shock to leveraged ETF investor demand. Specifically, the component ssi is the realization of a speculative demand shock. Both ω and ssi are drawn from mean zero distributions with variances σ_ω^2 and σ_{ssi}^2 respectively. The correlation between ω and ssi is,

$$\rho \equiv \mathrm{CORR}(\omega, ssi), \tag{IA7}$$

with $\rho \in [-1, 1]$.

ETF j's NAV at $t = 1$ is given by,

$$\pi_{1,j} = \Omega + (\omega + \varphi ssi)\,(1 - 2 \times \mathbb{1}_{j=S}), \tag{IA8}$$

in which $\varphi \in [0, 1)$ measures the extent to which ssi also affects the value of the underlying derivative security. Assuming $\varphi \in [0, 1)$ is consistent with my identifying assumption used in the empirical analysis; leveraged ETF share demand is *relatively* more sensitive to gambling-like, uninformed demand shocks than the underlying derivative security demand. Note, the $t = 1$ ETF premium for the leveraged-long ETF is,

$$\psi_{1,L} = (1 - \varphi) ssi, \tag{IA9}$$

and the $t = 1$ ETF premium for the leveraged-short ETF is,

$$\psi_{1,S} = -(1 - \varphi) ssi. \tag{IA10}$$

In other words, because the leveraged ETFs' shares and the underlying derivative security have different sensitivities to speculative demand shocks, the realization of a nonzero ssi

gives rise to relative mispricing.

At $t = 2$, authorized participants trade against $\psi_{1,L}$ and $\psi_{1,S}$. Their trades have price impact and the $t = 2$ price of ETF j is given by,

$$p_{2,j}(\Delta_j) = \beta_j - \eta_j q_{0,j} + (\omega + ssi)\,(1 - 2 \times \mathbb{1}_{j=S}) - \eta_j \Delta_j, \tag{IA11}$$

which is a function of Δ_j and differs from $p_{1,j}$ only by the price impact of the trades $-\eta_j \Delta_j$ (shorting ETF j's shares or purchasing ETF j's shares). Furthermore, ETF j's NAV at $t = 2$ is given by,

$$\pi_{2,j}(\Delta_L, \Delta_S) = \Omega + (\omega + \varphi ssi + \lambda\,(\Delta_L - \Delta_S))\,(1 - 2 \times \mathbb{1}_{j=S}), \tag{IA12}$$

which is a function of both Δ_L and Δ_S and in which $\lambda \geq 0$ represents the price impact that hedging via the derivative contract has.[2] Note, $\pi_{2,j}$ differs from $\pi_{1,j}$ only by $\lambda\,(\Delta_L - \Delta_S)\,(1 - 2 \times \mathbb{1}_{j=S})$.

Given the model's timing and the evolution of prices, I now solve for each authorized participant's optimal creation/redemption activity in the two ETFs. At the beginning of $t = 2$, each authorized participant i chooses a length $\delta_{L,i} \in \mathbb{R}$ of shares to create or redeem in the leveraged-long ETF and a length $\delta_{S,i} \in \mathbb{R}$ of shares to create or redeem in the leveraged-short ETF. Each authorized participant is concerned with

[2] Note, because the derivative market is perfectly competitive and frictionless, any price impact (i.e., $\lambda \neq 0$) is also reflected in the value of the underlying index χ. For example, the price impact from authorized participant trades may be transmitted to the underlying index via counterparties hedging in the physical underlying.

maximizing arbitrage profits and he solves the following optimization,

$$
\begin{aligned}
&\max_{\delta_{L,i}\in\mathbb{R},\delta_{S,i}\in\mathbb{R}} \delta_{L,i}\big(p_{2,L}(\delta_{L,i}+\delta_{L,-i}) - \pi_{2,L}(\delta_{L,i}+\delta_{L,-i},\delta_{S,i}+\delta_{S,-i})\big) \\
&\qquad + \delta_{S,i}\big(p_{2,S}(\delta_{S,i}+\delta_{S,-i}) - \pi_{2,S}(\delta_{L,i}+\delta_{L,-i},\delta_{S,i}+\delta_{S,-i})\big), \tag{IA13}
\end{aligned}
$$

in which $\delta_{j,-i}$ denotes the trades of the other $N-1$ authorized participants. To provide the most natural and intuitive explanation of the model, hereafter in the main prose I focus on the special case in which $\eta_L = \eta_S \equiv \eta$, and in the limit $N \to \infty$. Assuming $\eta_L = \eta_S \equiv \eta$ implies that investor demands in the leveraged-long ETF and leveraged-short ETF are equally sensitive to changes in shares outstanding. Considering the limiting case in which $N \to \infty$ focuses on perfect competition in the ETF primary market. While I focus on this special case in the prose, all supporting proofs are performed under the general case.

Lemma IA1. *The aggregate trades of authorized participants are,*

$$\lim_{N\to\infty,\eta_L=\eta_S=\eta} \Delta_L^* = \frac{ssi(1-\varphi)}{2\lambda+\eta}, \tag{IA14}$$

$$\lim_{N\to\infty,\eta_L=\eta_S=\eta} \Delta_S^* = -\frac{ssi(1-\varphi)}{2\lambda+\eta}. \tag{IA15}$$

According to Lemma IA1, a positive speculative demand shock (i.e., $ssi > 0$) is associated with share creations in the leveraged-long ETF and share redemptions in the leveraged-short ETF. For a negative speculative demand shock, the opposite holds. Notably, the fundamental shock ω does not appear in Δ_L^* or in Δ_S^*. Furthermore, Lemma IA1 may be used to generate a measure SSI^* as,

$$SSI^* \equiv \lim_{N\to\infty,\eta_L=\eta_S=\eta} \Delta_L^* - \lim_{N\to\infty,\eta_L=\eta_S=\eta} \Delta_S^* \tag{IA16}$$

$$= \frac{2ssi(1-\varphi)}{2\lambda+\eta}. \tag{IA17}$$

As such, SSI^* in the model is akin to the Speculation Sentiment Index used in the empirical analysis.

I now proceed to solving for the equilibrium $t = 2$ ETF share prices and NAVs.

Lemma IA2. *The equilibrium $t = 2$ ETF share prices and NAVs are given by,*

$$\lim_{N\to\infty,\eta_L=\eta_S=\eta} p_{2,L}(\Delta_L^*) = \lim_{N\to\infty,\eta_L=\eta_S=\eta} \pi_{2,L}(\Delta_L^*, \Delta_S^*) = \Omega + \omega + ssi\frac{2\lambda+\eta\varphi}{2\lambda+\eta}, \tag{IA18}$$

$$\lim_{N\to\infty,\eta_L=\eta_S=\eta} p_{2,S}(\Delta_S^*) = \lim_{N\to\infty,\eta_L=\eta_S=\eta} \pi_{2,S}(\Delta_L^*, \Delta_S^*) = \Omega - \omega - ssi\frac{2\lambda+\eta\varphi}{2\lambda+\eta} \tag{IA19}$$

Furthermore, both $\psi_{2,L}$ and $\psi_{2,S}$ equal zero.

Lemma IA2 shows that, with perfect competition, the equilibrium $t = 2$ ETF share prices are equal to their respective NAVs. That is, perfect competition among authorized participants is sufficiently strong to eliminate any relative mispricing between the ETF shares and their NAVs. The following lemma provides the $t =$ *long term* prices, which are equal to their latent fundamental values by assumption.

Lemma IA3. *The $t =$ long term ETF share prices and NAVs are given by,*

$$p_{long\ term,L}(\Delta_L^*) = \pi_{long\ term,L}(\Delta_L^*, \Delta_S^*) = \Omega + \omega, \tag{IA20}$$

$$p_{long\ term,S}(\Delta_S^*) = \pi_{long\ term,S}(\Delta_L^*, \Delta_S^*) = \Omega - \omega. \tag{IA21}$$

Using Lemma IA2 and Lemma IA3, price changes for the ETFs and NAVs may be computed. I focus on two periods: i) the price changes from $t = 0$ to $t = 2$ and ii) the price changes from $t = 2$ to $t =$ *long term*.[3] I denote a price change by α and I classify the change between $t = 0$ and $t = 2$ as the contemporaneous change (denoted with superscript C) and the change between $t = 2$ and $t =$ *long term* as the future change (denoted with the superscript F). Furthermore, since I am focusing on the limit $N \to \infty$, the price changes for the ETFs' shares and their respective NAVs are equal.

Lemma IA4. *The contemporaneous price changes are given by,*

$$\lim_{N\to\infty, \eta_L=\eta_S=\eta} \alpha_L^C = \omega + ssi\frac{2\lambda + \eta\varphi}{2\lambda + \eta}, \tag{IA22}$$

$$\lim_{N\to\infty, \eta_L=\eta_S=\eta} \alpha_S^C = -\omega - ssi\frac{2\lambda + \eta\varphi}{2\lambda + \eta}. \tag{IA23}$$

[3] I do not focus on the price changes between $t = 0$ and $t = 1$ and between $t = 1$ and $t = 2$, as the time it takes for authorized participants to exploit relative mispricing may be quick and unobservable to the empiricist.

The future price changes are given by,

$$\text{(IA24)} \qquad \lim_{N\to\infty,\eta_L=\eta_S=\eta} \alpha_L^F = -ssi\frac{2\lambda+\eta\varphi}{2\lambda+\eta},$$

$$\text{(IA25)} \qquad \lim_{N\to\infty,\eta_L=\eta_S=\eta} \alpha_S^F = ssi\frac{2\lambda+\eta\varphi}{2\lambda+\eta}.$$

I now explore the empirical implications from the preceding analysis. Specifically, I consider the sign of the slope coefficient from univariate regressions in which the independent variable is SSI^* and the dependent variable is either α_L^C or α_L^F (i.e., either the contemporaneous or future price change in the leveraged-long ETF). Considering α_S^C and α_S^F as dependent variables is redundant as slope coefficients are mechanically the negative of the coefficients using α_L^C or α_L^F. Furthermore, I note that while the dependent variable is the leveraged ETF share price change (or NAV change), the signs on the coefficients would be the same if one used the price change on the underlying index χ instead. As such, without loss of generality, I do not calculate or explicitly consider price changes for the benchmark index.[4]

Proposition IA1. *In a regression in which the independent variable is SSI^* and the dependent variable is contemporaneous price change α_L^C, the regression slope coefficient is given by,*

$$\text{(IA26)} \qquad b_{SSI}^C \equiv \frac{Cov(\alpha_L^C, SSI^*)}{Var(SSI^*)} = \rho\frac{\sigma_\omega}{2\sigma_{ssi}}\left(\frac{2\lambda+\eta}{1-\varphi}\right) + \left(\frac{2\lambda+\eta\varphi}{2(1-\varphi)}\right).$$

[4]For example, ω reflects the fundamental price change in the leveraged ETF share price and NAV. For the nonleveraged benchmark index, the magnitude of the fundamental price change would be attenuated by one half, all else equal.

b^C_{SSI} *is positive valued if* $\rho > \overline{\rho}$, *is negative valued if* $\rho < \overline{\rho}$, *and is equal to zero if* $\rho = \overline{\rho}$ *in which,*

$$\overline{\rho} \equiv -\frac{\sigma_{ssi}(2\lambda + \eta\varphi)}{\sigma_\omega(2\lambda + \eta)} < 0. \quad \text{(IA27)}$$

In a regression in which the independent variable is SSI^* *and the dependent variable is future price change* α^F_L, *the regression slope coefficient is given by,*

$$b^F_{SSI} \equiv \frac{Cov(\alpha^F_L, SSI^*)}{Var(SSI^*)} = -\frac{2\lambda + \eta\varphi}{2(1-\varphi)}. \quad \text{(IA28)}$$

b^F_{SSI} *is weakly negative valued and strictly negative valued if* $\varphi > 0$.

Corollary IA1. *If* Δ^*_L *is used in place of* SSI^* *in the regressions, the signs on the slope coefficients are given by,*

$$Sign(b^C_{\Delta_L}) = Sign(b^C_{SSI}), \quad \text{(IA29)}$$

$$Sign(b^F_{\Delta_L}) = Sign(b^F_{SSI}). \quad \text{(IA30)}$$

If Δ^*_S *is used in place of* SSI^* *in the regressions, the signs on the slope coefficients are given by,*

$$Sign(b^C_{\Delta_S}) = -1 \times Sign(b^C_{SSI}), \quad \text{(IA31)}$$

$$Sign(b^F_{\Delta_S}) = -1 \times Sign(b^F_{SSI}). \quad \text{(IA32)}$$

According to Proposition IA1, SSI^* negatively predicts *future* price changes. That is, a positive value of SSI^* is associated with negative future price changes in the benchmark index χ. Furthermore, Corollary IA1 shows that share changes in the leveraged-long ETF also have a negative relation with future price changes in the benchmark index while share changes in the leveraged-short ETF have a positive relation. These insights yield the following remark.

Remark IA1. *In empirical tests, both* SSI^* *and* Δ_L^* *should negatively predict future index returns.* Δ_S^* *should positively predict future index returns.*

Proposition IA1 also shows that the relation between SSI^* and contemporaneous price changes is equivocal and depends on the correlation between the fundamental demand shock ω and the speculative demand shock ssi. If $\rho > 0$, the relation between ω and ssi is best characterized as *extrapolative*; speculative demand shocks amplify fundamental news. If $\rho < 0$, the relation between ω and ssi is best characterized as *contrarian*; speculative demand shocks attenuate fundamental news. The threshold condition in Proposition IA1 implies that if speculative demand shocks are sufficiently contrarian (i.e., $\rho < \overline{\rho} < 0$), then the regression coefficient b_{SSI}^C is negative valued. As such, observing a negative relation between SSI^* and contemporaneous price changes in data is evidence that ρ is negative valued. Conversely, observing a positive relation between SSI^* and contemporaneous price changes is not indicative of whether ρ is negative valued or positive valued.

Remark IA2. *In empirical tests, a negative relation between* SSI^* *and contemporaneous returns is evidence that speculative demand shocks are contrarian.*

Note, Remark IA2 suggests that the negative relation between contemporaneous returns and SSI documented in Table 3 is sufficient evidence to confirm that speculative demand shocks themselves are contrarian (i.e., speculative sentiment bets against fundamental news). While both the empirical evidence throughout the paper and the model are consistent with this interpretation, a few disclaimers are in order. First, the model is highly stylized as it is intended to provide guiding intuition. As such, the model does not consider additional market frictions which could confound classifying demand shocks as either contrarian or extrapolative. Second, in the empirical analysis, while the *measure* of SSI is contrarian, there is no means to cleanly show that the nonfundamental demand measured by SSI itself is contrarian . For example, SSI could be comprised by two pieces:

$$\text{SSI} = \text{SENTIMENT\_SHOCKS} + \text{REBALANCING\_DEMAND}.$$

While the analysis in Section V.A shows that the return predictability is not driven by REBALANCING_DEMAND and instead is driven by SENTIMENT_SHOCKS, it is not possible to cleanly disentangle the two for contemporaneous returns. Specifically, REBALANCING_DEMAND is highly correlated (perhaps collinear) with contemporaneous returns. As such, the negative relation between SSI and contemporaneous returns may be driven by REBALANCING_DEMAND as opposed to SENTIMENT_SHOCKS.

While the model assumes that there are no limits to arbitrage (e.g., transaction costs), in practice arbitrage capital is limited and arbitrageurs face bid-ask spreads. As such, to the extent that Δ_L^* and Δ_S^* are measured with noise and the noise between the two is positively correlated, SSI^* may be a cleaner measure of speculative demand shocks than either Δ_L^* or Δ_S^* alone. Specifically, since SSI^* is the difference between Δ_L^* and Δ_S^*, any positively correlated noise on both share changes will be attenuated.

Remark IA3. *If Δ_L^* and Δ_S^* are measured with noise and the noise terms are positively correlated (e.g., variation in the cost of arbitrage capital), SSI^* provides a cleaner measure of speculative demand shocks.*

Finally, it is worth noting that the leveraged ETF market is relatively small as compared to the broad equity market. As such, the price impact of authorized participant trades on the underlying assets may be minimal (i.e., $\lambda \approx 0$). Nevertheless, in the limit $\lambda \to 0$,

$$\text{(IA33)} \qquad \lim_{\lambda \to 0} b_{SSI}^C \equiv \frac{\text{Cov}(\alpha_L^C, SSI^*)}{\text{Var}(SSI^*)} = \rho \frac{\sigma_\omega}{2\sigma_{ssi}} \left(\frac{\eta}{1-\varphi} \right) + \left(\frac{\eta\varphi}{2(1-\varphi)} \right),$$

$$\text{(IA34)} \qquad \lim_{\lambda \to 0} b_{SSI}^F \equiv \frac{\text{Cov}(\alpha_L^F, SSI^*)}{\text{Var}(SSI^*)} = -\frac{\varphi\eta}{2(1-\varphi)}.$$

A sufficient condition for both b_{SSI}^C and b_{SSI}^F to be nonzero valued is that $\varphi > 0$ (i.e., speculative demand shocks are market-wide). The following remark highlights this insight.

Remark IA4. *Even if authorized participant trades in the leveraged ETF market have no price impact on the underlying assets, SSI^* provides a measure of market-wide speculative demand shocks.*

IA.2.1 Supporting Proofs

Proof of Lemma IA1: Denote each authorized participant's objective function, given the trades of the other $N-1$ authorized participants, as $\Pi(\delta_{L,i}, \delta_{S,i})$. $\Pi(\delta_{L,i}, \delta_{S,i})$ is explicitly given by,

$$\begin{aligned}\Pi(\delta_{L,i}, \delta_{S,i}) =& \delta_{L,i}\left((1-\varphi)ssi - \eta_L(\delta_{L,i}+\delta_{L,-i}) - \lambda(\delta_{L,i}+\delta_{L,-i}-\delta_{S,i}-\delta_{S,-i})\right) \\ &+ \delta_{S,i}\left(-(1-\varphi)ssi - \eta_S(\delta_{S,i}+\delta_{S,-i}) + \lambda(\delta_{L,i}+\delta_{L,-i}-\delta_{S,i}-\delta_{S,-i})\right). \end{aligned} \tag{IA35}$$

It is straightforward that $\Pi(\delta_{L,i}, \delta_{S,i})$ is concave in $\{\delta_{L,i}, \delta_{S,i}\}$ because,

$$\begin{aligned}\frac{\partial^2\Pi}{\partial\delta_{L,i}^2} &= -2(\eta_L+\lambda) \\ &< 0, \end{aligned} \tag{IA36}$$

$$\begin{aligned}\frac{\partial^2\Pi}{\partial\delta_{S,i}^2} &= -2(\eta_S+\lambda) \\ &< 0, \end{aligned} \tag{IA37}$$

$$\begin{aligned}\frac{\partial^2\Pi}{\partial\delta_{L,i}\partial\delta_{S,i}} &= 2\lambda \\ &\geq 0. \end{aligned} \tag{IA38}$$

Therefore, first-order conditions are necessary and sufficient to solve for each authorized participants optimal choices $\{\delta_{L,i}^*, \delta_{S,i}^*\}$.

The first-order conditions yields a system of two equations with two unknowns.

Solving the system yields,

$$\delta_{L,i} = \frac{-\lambda\eta_L\delta_{L,-i} - \eta_S(\eta_L + \lambda)\delta_{L,-i} + \eta_S ssi(1-\varphi)}{2((\eta_L + \eta_S)\lambda + \eta_L\eta_S)}, \tag{IA39}$$

$$\delta_{S,i} = \frac{-\lambda\eta_S\delta_{S,-i} - \eta_L(\eta_S + \lambda)\delta_{S,-i} - \eta_L ssi(1-\varphi)}{2((\eta_L + \eta_S)\lambda + \eta_L\eta_S)}. \tag{IA40}$$

Next, the symmetric solution may be solved for by substituting $(N-1)\delta_{j,i}$ in place of $\delta_{j,-i}$ into equation (IA39) and equation (IA40) and then solving for $\delta_{L,i}$ and $\delta_{S,i}$. Doing so yields,

$$\delta^*_{L,i} = \frac{\eta_S ssi(1-\varphi)}{(N+1)((\eta_L + \eta_S)\lambda + \eta_L\eta_S)}, \tag{IA41}$$

$$\delta^*_{S,i} = \frac{-\eta_L ssi(1-\varphi)}{(N+1)((\eta_L + \eta_S)\lambda + \eta_L\eta_S)}. \tag{IA42}$$

Therefore, the aggregate trades are,

$$\Delta^*_L = \frac{N\eta_S ssi(1-\varphi)}{(N+1)((\eta_L + \eta_S)\lambda + \eta_L\eta_S)}, \tag{IA43}$$

$$\Delta^*_S = \frac{-N\eta_L ssi(1-\varphi)}{(N+1)((\eta_L + \eta_S)\lambda + \eta_L\eta_S)}. \tag{IA44}$$

Finally, in the special case in which $\eta_L = \eta_S \equiv \eta$ and in the limit $N \to \infty$, the aggregate trades simplify to,

$$\Delta^*_L = \frac{ssi(1-\varphi)}{2\lambda + \eta}, \tag{IA45}$$

$$\Delta^*_S = -\frac{ssi(1-\varphi)}{2\lambda + \eta}. \tag{IA46}$$

■

Proof of Lemma IA2: Using the results from Lemma IA1 and equation (IA5), the $t = 2$ share prices are given by,

$$p_{2,L}(\Delta_L^*) = \Omega + \omega + ssi - \eta_L \left(\frac{N\eta_S ssi(1-\varphi)}{(N+1)((\eta_L+\eta_S)\lambda + \eta_L\eta_S)} \right), \tag{IA47}$$

$$p_{2,S}(\Delta_S^*) = \Omega - \omega - ssi - \eta_S \left(\frac{-N\eta_L ssi(1-\varphi)}{(N+1)((\eta_L+\eta_S)\lambda + \eta_L\eta_S)} \right). \tag{IA48}$$

Additionally, the $t = 2$ NAVs are given by,

$$\pi_{2,L}(\Delta_L^*, \Delta_S^*) = \Omega + \left(\omega + \varphi ssi + \lambda \left(\frac{N(\eta_L+\eta_S)ssi(1-\varphi)}{(N+1)((\eta_L+\eta_S)\lambda + \eta_L\eta_S)} \right) \right), \tag{IA49}$$

$$\pi_{2,S}(\Delta_L^*, \Delta_S^*) = \Omega - \left(\omega + \varphi ssi + \lambda \left(\frac{N(\eta_L+\eta_S)ssi(1-\varphi)}{(N+1)((\eta_L+\eta_S)\lambda + \eta_L\eta_S)} \right) \right). \tag{IA50}$$

Finally, in the special case in which $\eta_L = \eta_S \equiv \eta$ and in the limit $N \to \infty$, the $t = 2$ ETF share prices and NAVs simplify to,

$$\lim_{N\to\infty,\eta_L=\eta_S=\eta} p_{2,L}(\Delta_L^*) = \lim_{N\to\infty,\eta_L=\eta_S=\eta} \pi_{2,L}(\Delta_L^*, \Delta_S^*) = \Omega + \omega + ssi - \frac{ssi(1-\varphi)\eta}{2\lambda+\eta}, \tag{IA51}$$

$$\lim_{N\to\infty,\eta_L=\eta_S=\eta} p_{2,S}(\Delta_S^*) = \lim_{N\to\infty,\eta_L=\eta_S=\eta} \pi_{2,S}(\Delta_L^*, \Delta_S^*) = \Omega - \omega - ssi + \frac{ssi(1-\varphi)\eta}{2\lambda+\eta}. \tag{IA52}$$

■

Proof of Lemma IA3:

The prices, by assumption, return to their latent fundamental values.

■

Proof of Lemma IA4:

The proof follows directly from the main text.

■

Proof of Proposition IA1 and Corollary IA1: First, I provide several results that are subsequently used to derive the regression slope coefficients. First, the variances of $\lim_{N\to\infty,\eta_L=\eta_S=\eta} \Delta_L^*$ and $\lim_{N\to\infty,\eta_L=\eta_S=\eta} \Delta_S^*$ are both given by,

$$\text{Var}(\Delta_L^*) = \text{Var}(\Delta_S^*) = \left(\frac{1-\varphi}{2\lambda+\eta}\right)^2 \sigma_{ssi}^2, \tag{IA53}$$

and the variance of SSI^* is given by,

$$\text{Var}(SSI^*) = \left(\frac{2(1-\varphi)}{2\lambda+\eta}\right)^2 \sigma_{ssi}^2. \tag{IA54}$$

Second, the covariance between $\lim_{N\to\infty,\eta_L=\eta_S=\eta} \Delta_L^*$ and α_L^C is given by,

$$\text{Cov}(\alpha_L^C, \Delta_L^*) = \rho\sigma_\omega\sigma_{ssi}\left(\frac{1-\varphi}{2\lambda+\eta}\right) + \sigma_{ssi}^2\left(\frac{(2\lambda+\eta\varphi)(1-\varphi)}{(2\lambda+\eta)^2}\right), \tag{IA55}$$

and the covariance between $\lim_{N\to\infty,\eta_L=\eta_S=\eta} \Delta_S^*$ and α_L^C is given by,

$$\text{Cov}(\alpha_L^C, \Delta_S^*) = -\rho\sigma_\omega\sigma_{ssi}\left(\frac{1-\varphi}{2\lambda+\eta}\right) - \sigma_{ssi}^2\left(\frac{(2\lambda+\eta\varphi)(1-\varphi)}{(2\lambda+\eta)^2}\right). \tag{IA56}$$

The covariance between SSI^* and α_L^C is given by,

$$\text{(IA57)} \qquad \text{Cov}(\alpha_L^C, SSI^*) = \rho\sigma_\omega\sigma_{ssi}\left(\frac{2(1-\varphi)}{2\lambda+\eta}\right) + \sigma_{ssi}^2\left(\frac{2(2\lambda+\eta\varphi)(1-\varphi)}{(2\lambda+\eta)^2}\right).$$

Third, the covariance between $\lim\limits_{N\to\infty,\eta_L=\eta_S=\eta} \Delta_L^*$ and α_L^F is given by,

$$\text{(IA58)} \qquad \text{Cov}(\alpha_L^F, \Delta_L^*) = -\sigma_{ssi}^2\left(\frac{(2\lambda+\eta\varphi)(1-\varphi)}{(2\lambda+\eta)^2}\right),$$

and the covariance between $\lim\limits_{N\to\infty,\eta_L=\eta_S=\eta} \Delta_S^*$ and α_L^F is given by,

$$\text{(IA59)} \qquad \text{Cov}(\alpha_L^F, \Delta_S^*) = \sigma_{ssi}^2\left(\frac{(2\lambda+\eta\varphi)(1-\varphi)}{(2\lambda+\eta)^2}\right).$$

The covariance between SSI^* and α_L^F is given by,

$$\text{(IA60)} \qquad \text{Cov}(\alpha_L^F, SSI^*) = -\sigma_{ssi}^2\left(\frac{2(2\lambda+\eta\varphi)(1-\varphi)}{(2\lambda+\eta)^2}\right).$$

Using the preceding results, the slope coefficients using contemporaneous price changes as the dependent variable are,

$$\text{(IA61)} \qquad \frac{\text{Cov}(\alpha_L^C, \Delta_L^*)}{\text{Var}(\Delta_L^*)} = \rho\frac{\sigma_\omega}{\sigma_{ssi}}\left(\frac{2\lambda+\eta}{1-\varphi}\right) + \left(\frac{2\lambda+\eta\varphi}{1-\varphi}\right),$$

$$\text{(IA62)} \qquad \frac{\text{Cov}(\alpha_L^C, \Delta_S^*)}{\text{Var}(\Delta_S^*)} = -\rho\frac{\sigma_\omega}{\sigma_{ssi}}\left(\frac{2\lambda+\eta}{1-\varphi}\right) - \left(\frac{2\lambda+\eta\varphi}{1-\varphi}\right),$$

$$\text{(IA63)} \qquad \frac{\text{Cov}(\alpha_L^C, SSI^*)}{\text{Var}(SSI^*)} = \rho\frac{\sigma_\omega}{2\sigma_{ssi}}\left(\frac{2\lambda+\eta}{1-\varphi}\right) + \left(\frac{2\lambda+\eta\varphi}{2(1-\varphi)}\right).$$

The signs on the preceding coefficients are determined by ρ. Define,

$$\text{(IA64)} \qquad \overline{\rho} \equiv -\frac{\sigma_{ssi}(2\lambda + \eta\varphi)}{\sigma_\omega(2\lambda + \eta)}.$$

Using $\overline{\rho}$, the signs are given by,

$$\text{(IA65)} \qquad \text{Sign}\left(\frac{\text{Cov}(\alpha_L^C, \Delta_L^*)}{\text{Var}(\Delta_L^*)}\right) = \begin{cases} + & \text{if } \rho > \overline{\rho} \\ 0 & \text{if } \rho = \overline{\rho} \\ - & \text{if } \rho < \overline{\rho}, \end{cases}$$

$$\text{(IA66)} \qquad \text{Sign}\left(\frac{\text{Cov}(\alpha_L^C, \Delta_S^*)}{\text{Var}(\Delta_S^*)}\right) = \begin{cases} - & \text{if } \rho > \overline{\rho} \\ 0 & \text{if } \rho = \overline{\rho} \\ + & \text{if } \rho < \overline{\rho}, \end{cases}$$

$$\text{(IA67)} \qquad \text{Sign}\left(\frac{\text{Cov}(\alpha_L^C, SSI^*)}{\text{Var}(SSI^*)}\right) = \begin{cases} + & \text{if } \rho > \overline{\rho} \\ 0 & \text{if } \rho = \overline{\rho} \\ - & \text{if } \rho < \overline{\rho}. \end{cases}$$

Also using the preceding results, the slope coefficients using future price changes as the dependent variable are,

$$\text{(IA68)} \qquad \frac{\text{Cov}(\alpha_L^F, \Delta_L^*)}{\text{Var}(\Delta_L^*)} = -\frac{2\lambda + \eta\varphi}{1-\varphi},$$

$$\text{(IA69)} \qquad \frac{\text{Cov}(\alpha_L^F, \Delta_S^*)}{\text{Var}(\Delta_S^*)} = \frac{2\lambda + \eta\varphi}{1-\varphi},$$

$$\text{(IA70)} \qquad \frac{\text{Cov}(\alpha_L^F, SSI^*)}{\text{Var}(SSI^*)} = -\frac{2\lambda + \eta\varphi}{2(1-\varphi)}.$$

The signs on the preceding coefficients given by,

$$\text{(IA71)} \qquad \text{Sign}\left(\frac{\text{Cov}(\alpha_L^F, \Delta_L^*)}{\text{Var}(\Delta_L^*)}\right) = -,$$

$$\text{(IA72)} \qquad \text{Sign}\left(\frac{\text{Cov}(\alpha_L^F, \Delta_S^*)}{\text{Var}(\Delta_S^*)}\right) = +,$$

$$\text{(IA73)} \qquad \text{Sign}\left(\frac{\text{Cov}(\alpha_L^F, SSI^*)}{\text{Var}(SSI^*)}\right) = -.$$

■

IA.3 Pitfalls With Higher Frequency Measures

A clear advantage of SSI relative to other sentiment measures is the frequency at which it may be calculated; ETF share changes are reported on a daily basis. Therefore, one can construct a speculation sentiment measure at the daily frequency as easily as one can construct it at the monthly frequency. However, one should be cautious in using higher frequency (e.g., daily or weekly) measures constructed from ETF share changes.

First, Staer (2017) shows that ETFs often report using $T+1$ accounting meaning that shares outstanding (and share changes) are reported with a one day lag, but that the lag is time-varying and may at times be T accounting. Furthermore, changes in reporting lag are not public. This implies that daily share change data may be one-day stale on some dates and not stale on others. Second, Evans, Moussawi, Pagano and Sedunov (2019) describes how APs can strategically delay the creation of new shares until $T + 6$. By doing so, APs avoid costs associated with short-selling and it also allows APs to strategically time return reversals (since the authorized participants are essentially engaging in a naked short position). Therefore, given potential accounting inconsistencies and strategic delay in creating new ETF shares, observed share change activity on a given date might be related to market activities from several trading days prior.

IA.4 Small Sample Parametric Bootstrap

Stambaugh (1999) highlights potential biases in predictive regressions, that is, the OLS estimator's small sample properties violate standard regression assumptions. As such, t-statistics that do not account for this bias may be inflated and give a false sense of statistical significance. Therefore, to ensure that the coefficients on SSI_t in Table 3 are statistically significant, I correct for a potential Stambaugh bias in this ancillary analysis. Specifically, I compute p-values using a small sample parametric bootstrap detailed below.

Let r_t be the return on the benchmark index in period t and SSI_t be the value of the Speculation Sentiment Index at the end of period t. The univariate predictive regressions reported in Panel B of Table 3 are of the form,

$$r_{t+1} = a + \beta \text{SSI}_t + \epsilon_{t+1}. \tag{IA74}$$

I estimate the coefficients using OLS and the t-statistics are computed as GMM corrected standard errors (with equal weighting). Denote the t-statistic for β as τ. After obtaining the coefficient estimates and t-statistics, I estimate the small-sample distribution of the t-statistics under the null hypothesis of no predictability. To obtain the distribution, I perform the following bootstrap procedure:

i) I estimate the restricted VAR,

$$\begin{bmatrix} r_{t+1} \\ \text{SSI}_{t+1} \end{bmatrix} = A + \begin{bmatrix} 0 & 0 \\ 0 & \phi \end{bmatrix} \begin{bmatrix} r_t \\ \text{SSI}_t \end{bmatrix} + \epsilon_{t+1}, \tag{IA75}$$

and keep the residuals ϵ_{t+1}.

ii) For each bootstrap simulation I,

(a) Initialize $\begin{bmatrix} r_0 \\ \text{SSI}_0 \end{bmatrix}$ to their unconditional means.

(b) For $t = 1$ through $t = \overline{T}$, let,

$$\begin{bmatrix} r_{t+1} \\ \text{SSI}_{t+1} \end{bmatrix} = A + \begin{bmatrix} 0 & 0 \\ 0 & \phi \end{bmatrix} \begin{bmatrix} r_t \\ \text{SSI}_t \end{bmatrix} + e_{t+1}, \tag{IA76}$$

in which e_{t+1} is a random draw (with replacement) from the residuals ϵ_{t+1} recovered in step i).

(c) I throw away the "burn-in" initial data and keep the last T observations corresponding to the length of the original data sample. I then estimate a coefficient $\hat{\beta}$ and corresponding t-statistic $\hat{\tau}$ using the simulated data.

iii) I use the bootstrap distribution of the $\hat{\tau}$ to get a p-value for the actual t-statistic τ.

In total, I use 1,000,000 bootstrap simulations and in each simulation $\overline{T}$ is equal to 300.[5] In addition to using the small sample parametric bootstrap to obtain p-values for the regression coefficients, I also use the simulated data to estimate the bias in the regression coefficients and to assess the statistical significance of realized adjusted R^2's. See Panel B of Table 3. This procedure is also utilized in Section V.D to assess the statistical significance of the out-of-sample R^2's reported in Table 8. Note, for that analysis, I utilize 10,000 bootstrap simulations.

IA.5 Daily Returns and Daily SSI

Using monthly returns and monthly values of SSI, the results in Section IV show a negative relation between SSI and contemporaneous returns. One may be concerned that the results are an artifact of timing within the month and that, at higher frequencies, no contrarian relation between SSI and returns exists. To address this, I construct SSI_t^{daily} as

[5]To compute the GMM corrected standard errors I use John Cochrane's olsgmm.m Matlab function. Furthermore, I am indebted to Shri Santosh for his comments on this analysis.

the daily Speculation Sentiment Index and I examine whether or not lagged daily returns predict the SSI_t^{daily} value. In Section IA.3 of this Internet Appendix, I highlight that there are potential pitfalls using daily data (such as stale data). Thus, I address the possibility of stale share change concerns by including five daily lagged returns. The predictive regression used with SSI_t^{daily} is,

$$\text{SSI}_t^{daily} = a + \beta_{0d} r_t + \beta_{1d} r_{t-1d} + \beta_{2d} r_{t-2d} + \beta_{3d} r_{t-3d} + \beta_{4d} r_{t-4d} + \beta_{5d} r_{t-5d} + \epsilon_t, \tag{IA77}$$

in which SSI_t^{daily} is the daily Speculation Sentiment Index value, a is the regression intercept, r_{t-id} is the index daily return i days before day t, β_{id} is the estimated coefficient on r_{t-id}, and ϵ_t is the error term. The results are reported in Table IA1. Regressions 1-3 report the results using the full sample and Regressions 4-6 report the results using the post-2009 sample. Regressions 1 and 4 use the CRSP equal weighted index, Regressions 2 and 5 use the CRSP value weighted index, and Regressions 3 and 6 use the S&P 500 index.

Table IA1 shows that daily returns are strong predictors of daily SSI; the contemporaneous daily return and each of the previous five trading days' returns carry a *negative* coefficient and all coefficients are statistically significant. In other words, daily SSI is contrarian as it is negatively related to recent market returns, confirming the insight from Section IV.

Table IA1: Predicting SSI^{daily} with Daily Returns

The daily Speculation Sentiment Index value is regressed on the contemporaneous daily return and five lagged daily index returns for the CRSP equal weighted index, CRSP value weighted index, or S&P 500 index monthly index: $\text{SSI}_t^{daily} = a + \beta_{0d} r_t + \beta_{1d} r_{t-1d} + \beta_{2d} r_{t-2d} + \beta_{3d} r_{t-3d} + \beta_{4d} r_{t-4d} + \beta_{5d} r_{t-5d} + \epsilon_t$ in which SSI_t^{daily} is the daily Speculation Sentiment Index value, a is the regression intercept, r_{t-id} is the index daily return i days before day t, β_{id} is the estimated coefficient on r_{t-id}, and ϵ_t is the error term. In Regressions 1-3, the sample returns run from Nov. 2006 through Dec. 2019. In Regressions 4-6, the sample returns run from Jan. 2010 through Dec. 2019. SSI_t^{daily} is constructed with daily share change data from ProShares. White standard errors are used to account for heteroscedasticity and t-statistics are reported, in parenthesis, below each estimated coefficient. All variables, except for returns, are standardized.

Predicting Daily SSI with Daily Returns

	Full Sample			Post-2009		
	1 EW	2 VW	3 SP500	4 EW	5 VW	6 SP500
α	0.03**	0.03**	0.03*	0.01	0.01	0.01
	(2.06)	(2.04)	(1.82)	(0.58)	(0.85)	(0.75)
r_t	-0.19***	-0.18***	-0.18***	-0.09***	-0.09***	-0.08***
	(-9.31)	(-9.18)	(-8.96)	(-4.17)	(-4.52)	(-4.43)
r_{t-1}	-0.21***	-0.22***	-0.23***	-0.13***	-0.13***	-0.13***
	(-10.29)	(-11.02)	(-11.00)	(-7.82)	(-8.37)	(-8.25)
r_{t-2}	-0.14***	-0.16***	-0.17***	-0.11***	-0.11***	-0.11***
	(-6.95)	(-8.12)	(-8.27)	(-6.16)	(-6.79)	(-6.69)
r_{t-3}	-0.06***	-0.09***	-0.10***	-0.06***	-0.08***	-0.08***
	(-3.28)	(-4.86)	(-5.15)	(-3.71)	(-5.00)	(-5.04)
r_{t-4}	-0.06***	-0.08***	-0.09***	-0.03*	-0.04***	-0.05***
	(-3.16)	(-4.55)	(-4.84)	(-1.90)	(-2.99)	(-3.13)
r_{t-5}	-0.07***	-0.08***	-0.08***	-0.03**	-0.04**	-0.04**
	(-4.05)	(-4.57)	(-4.56)	(-1.97)	(-2.51)	(-2.50)
Adj. R^2	0.15	0.16	0.15	0.07	0.07	0.07
N	3344	3344	3344	2516	2516	2516

Note: $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$

IA.6 SSI as Rational Rebalancing

It is possible that SSI measures rational trading. To see this, consider an investor that desires a particular portfolio on the CML and requires leverage to achieve the portfolio. Because leveraged ETFs provide daily magnified exposure that does not compound, the investor must rebalance her portfolio daily to retain the target leverage quantity. Specifically, suppose the investor begins with one dollar of wealth and she desires a leverage quantity $m \in \{-3, -2, 2, 3\}$ and uses a leveraged ETF that provides daily $m\times$ exposure. Over any two consecutive days, the investor's objective is to achieve,

$$m\left(\prod_{i=1}^{2}(1+r_i)-1\right), \tag{IA78}$$

however, a buy-and-hold strategy with a leveraged ETF share yields,

$$\prod_{i=1}^{2}(1+mr_i)-1. \tag{IA79}$$

This implies that the investor must rebalance to have notional exposure ω_1 at the end of day 1 such that the following equation is satisfied,

$$m\left((1+r_1)(1+E[r_2])-1\right) = \left((1+mr_1)\omega_1(1+mE[r_2])+(1-\omega_1)(1+mr_1)-1\right). \tag{IA80}$$

If $\omega_1 < 1$, the investor holds a fraction $(1 - \omega_1)$ of her wealth $(1 + mr_1)$ in cash (and earns a rate of return equal to zero). Conversely, if $\omega_1 > 1$, the investor borrows a fraction $(1 - \omega_1)$ of her wealth $(1 + mr_1)$ (at a cost equal to zero). Using equation (IA80), ω_1 (the daily rebalancing value) is given explicitly by,

(IA81)
$$\omega_1 \equiv \frac{1 + r_1}{1 + mr_1}.$$

Note, that the change in ω_1 with respect to a change in r_1 is given by,

(IA82)
$$\frac{\mathrm{d}\omega_1}{\mathrm{d}r_1} \equiv \frac{1 - m}{(1 + mr_1)^2},$$

which is negative valued if m is positive and is positive valued if m is negative. In other words, if an investor purchases a leveraged-long ETF ($m > 0$), she must sell shares if market returns are positive and buy shares if market returns are negative. Conversely, if an investor purchases a leveraged-short ETF ($m < 0$), she must buy shares if market returns are positive and sell shares if market returns are negative. This implies that rational rebalancing is mechanically contrarian.[6]

If one assumes that all rebalancing is accomplished via share creations, this implies that the daily share change is linear in ω_1. As such, I construct implied rebalancing demand as a control. Specifically, I calculate the daily implied rebalancing demand using the expression in equation (IA81) and the realized leveraged ETF returns. For a

[6] For a similar discussion, see Ivanov and Lenkey (2014).

leveraged-long ETF, the daily implied rebalancing demand is equal to,

$$\omega_t^{L,\text{imp}} \equiv \frac{1 + \frac{r_t^L}{m}}{1 + r_t^L}, \tag{IA83}$$

in which r_t^L is the leveraged-long ETF's daily return. For a leveraged-short ETF, the daily implied rebalancing demand is equal to,

$$\omega_t^{S,\text{imp}} \equiv \frac{1 + \frac{r_t^S}{m}}{1 + r_t^S}, \tag{IA84}$$

in which r_t^S is the leveraged-short ETF's daily return. I then aggregate monthly net implied rebalancing demand REBAL_t as,

$$\text{REBAL}_t = \sum_{i \in J} \prod_{\tau=1}^{T} \omega_{i,\tau}^{L,\text{imp}} - \sum_{i \in K} \prod_{\tau=1}^{T} \omega_{i,\tau}^{S,\text{imp}}, \tag{IA85}$$

in which J is the set of leveraged-long ETFs and K is the set of leveraged-short ETFs. The products of $\omega_{i,\tau}^{L,\text{imp}}$ and $\omega_{i,\tau}^{S,\text{imp}}$ from $\tau = 1$ to $t = T$ reflect the compounding of share change over month t's T days. Over the entire data sample of Oct. 2006 through Nov. 2019, REBAL_t and SSI_t have a correlation coefficient of 0.61.

Table IA2 replicates the analysis from Table 4 but with REBAL_t in the place of r_t. Similar to the results in Table 4 using r_t, REBAL_t has little to no predictive power in the univariate regressions reported in Panel A. In fact, in the post-2009 sample, the coefficients on REBAL_t are all positive valued and the coefficient in Regression 6 is marginally significant. Similarly, Panel B of Table IA2 reports results consistent with those in Panel B

Table IA2: Predictive Regressions with Implied Rebalancing and SSI

In Panel A, the future CRSP equal weighted, CRSP value weighted, or S&P 500 index monthly returns are regressed on the implied rebalancing level: $r_{t+1} = a + \beta \text{REBAL}_t + \epsilon_{t+1}$ in which r_{t+1} is the future index monthly return, REBAL_t is the implied rebalancing level from equation (IA85), β is the estimated coefficient on REBAL_t, and ϵ_{t+1} is the error term. In Panel B, the future CRSP equal weighted, CRSP value weighted, or S&P 500 index monthly returns are regressed on SSI and the implied rebalancing level: $r_{t+1} = a + \beta_{SSI}\text{SSI}_t + \beta_{rebal}\text{REBAL}_t + \epsilon_{t+1}$ in which r_{t+1} is the future index monthly return, SSI_t is the monthly value of SSI, REBAL_t is the implied rebalancing level, β_{SSI} is the estimated coefficient on SSI_t, β_{rebal} is the estimated coefficient on REBAL_t and ϵ_{t+1} is the error term. For both panels, in Regressions 1-3, the sample returns run from Nov. 2006 through Dec. 2019 and in Regressions 4-6, the sample returns run from Jan. 2010 through Dec. 2019. White standard errors are used to account for heteroscedasticity and t-statistics are reported, in parenthesis, below each estimated coefficient. All variables, except for returns, are standardized.

Panel A: Univariate Predictive Regressions r_{t+1}

	Full Sample			Post-2009		
	1	2	3	4	5	6
	EW	VW	SP500	EW	VW	SP500
REBAL_t	-0.87	-0.44	-0.36	0.54	0.70	0.70*
	(-1.43)	(-0.89)	(-0.78)	(1.08)	(1.61)	(1.68)
Adj. R^2	0.02	0.00	0.00	0.01	0.02	0.02
N	158	158	158	120	120	120

Panel B: Bivariate Predictive Regressions r_{t+1}

	1	2	3	4	5	6
SSI_t	-1.81***	-1.60***	-1.46**	-2.53***	-2.26***	-2.10***
	(-2.81)	(-2.70)	(-2.60)	(-3.35)	(-3.08)	(-2.93)
REBAL_t	0.24	0.54	0.54	1.24**	1.33***	1.29***
	(0.39)	(1.07)	(1.09)	(2.53)	(3.04)	(3.02)
Adj. R^2	0.10	0.08	0.07	0.10	0.11	0.10
N	158	158	158	120	120	120

Note: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

of Table 4; SSI becomes a stronger predictor of returns with the inclusion of implied rebalancing.

IA.7 Alternative SSI Specifications

For robustness, in this ancillary analysis, I consider several alternative constructions of SSI that maintain the measure's economic interpretation but control for possible omitted variables or provide additional evidence that the predictability results from Section IV are not Type I errors.

IA.7.1 Autocorrelation Adjusted Index

Naturally, one might be concerned about persistence in SSI and the possibility of a Stambaugh-bias (Stambaugh (1999)). In this ancillary analysis, I document the autocorrelation of SSI. Specifically, in Panel A of Table IA3, I report the results of a regression of SSI_t on five of its lagged values,

$$\text{SSI}_t = a + \beta_1 \text{SSI}_{t-1} + \beta_2 \text{SSI}_{t-2} + \beta_3 \text{SSI}_{t-3} + \beta_4 \text{SSI}_{t-4} + \beta_5 \text{SSI}_{t-5} + \epsilon_t. \tag{IA86}$$

The coefficient β_1 on the first lagged value SSI_{t-1} carries a coefficient of approximately 0.29 and is statistically significant at a 1% p-value threshold. Given serial correlation across months, for robustness, I provide an alternative index by estimating SSI as an AR(1) process,

$$\text{SSI}_t = a + \gamma \text{SSI}_{t-1} + \text{SSI}_t^{AR}, \tag{IA87}$$

in which a is a constant, γ is the AR(1) coefficient on SSI_{t-1}, and SSI_t^{AR} is the innovation to the series. I use the time series of SSI from Oct. 2006 through Dec. 2019 to estimate the AR(1) process. After estimating the parameters a and γ, the series of innovations are given by,

$$\text{SSI}^{AR} \equiv \{\text{SSI}_1^{AR}, \ldots, \text{SSI}_T^{AR}\}. \tag{IA88}$$

Panel B of Table IA3 provides the AR(1) estimation and Panel C of Table IA3 presents the results of the regression of SSI_t^{AR} on five of its lagged values,

$$\text{SSI}_t^{AR} = a + \beta_1 \text{SSI}_{t-1}^{AR} + \beta_2 \text{SSI}_{t-2}^{AR} + \beta_3 \text{SSI}_{t-3}^{AR} + \beta_4 \text{SSI}_{t-4}^{AR} + \beta_5 \text{SSI}_{t-5}^{AR} + \epsilon_t. \tag{IA89}$$

The results do not exhibit autocorrelation in SSI^{AR}.

I repeat much of the regression analysis from Section IV and Section V but use SSI^{AR} in place of SSI. Table IA4 provides the results. Regression 1 presents the univariate regression results while Regressions 2-17 present the bivariate regression results with controls. The results in Table IA4 are nearly identical to the results in Table 3, Table 4, and Table 7; the coefficient values and corresponding t-statistics are nearly identical for regressions using the CRSP equal weighted index, CRSP value weighted index, and the S&P 500 index. The adjusted R^2's are nearly identical as well. The analysis with SSI^{AR} provides evidence that the return predictability results are not due to persistence in the predictor variable.

Table IA3: Autocorrelation in SSI_t and SSI_t^{AR}

Panel A presents the results of the regression $\text{SSI}_t = a + \beta_1\text{SSI}_{t-1} + \beta_2\text{SSI}_{t-2} + \beta_3\text{SSI}_{t-3} + \beta_4\text{SSI}_{t-4} + \beta_5\text{SSI}_{t-5} + \epsilon_t$, in which SSI_t is defined in equation (2), a is the regression intercept and ϵ_t is the error term. Panel B presents the estimation of the AR(1) process governing SSI_t^{AR}. The AR(1) process is estimated using OLS. Panel C presents the results of the regression $\text{SSI}_t^{AR} = a + \beta_1\text{SSI}_{t-1}^{AR} + \beta_2\text{SSI}_{t-2}^{AR} + \beta_3\text{SSI}_{t-3}^{AR} + \beta_4\text{SSI}_{t-4}^{AR} + \beta_5\text{SSI}_{t-5}^{AR} + \epsilon_t$, in which SSI_t^{AR} is the Speculation Sentiment Index controlling for autocorrelation on date t, a is the regression intercept and ϵ_t is the error term. In all panels, t-statistics are reported, in parenthesis, below each estimated coefficient.

Panel A: SSI_t Regression with Lags	
Intercept	-0.03
	(-0.53)
SSI_{t-1}	0.29***
	(3.61)
SSI_{t-2}	-0.09
	(-1.03)
SSI_{t-3}	0.16*
	(1.94)
SSI_{t-4}	-0.03
	(-0.32)
SSI_{t-5}	0.04
	(0.47)
Adj. R^2	0.07
N	154

Panel B: AR(1) Estimation	
Intercept	-0.04
	(-0.81)
SSI_{t-1}	0.28***
	(3.62)
Adj. R^2	0.07
N	158

Panel C: SSI_t^{AR} Regression with Lags	
Intercept	0.01
	(0.18)
SSI_{t-1}^{AR}	0.01
	(0.18)
SSI_{t-2}^{AR}	-0.09
	(-1.11)
SSI_{t-3}^{AR}	0.14*
	(1.78)
SSI_{t-4}^{AR}	0.01
	(0.14)
SSI_{t-5}^{AR}	0.10
	(1.21)
Adj. R^2	0.00
N	153

Note: $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$

Table IA4: Return Predictability and SSI_t^{AR}

Regression 1 regresses the future CRSP equal weighted, CRSP value weighted, or S&P 500 index monthly returns on the Speculation Sentiment Index controlling for autocorrelation: $r_{t+1} = a + \beta \text{SSI}_t^{AR} + \epsilon_{t+1}$ in which r_{t+1} is the future index monthly return, SSI_t^{AR} is the Speculation Sentiment Index controlling for autocorrelation, β is the estimated coefficient on SSI_t^{AR}, and ϵ_{t+1} is the error term. Regressions 2-17 regress the future CRSP equal weighted, CRSP value weighted, or S&P 500 index monthly returns on the Speculation Sentiment Index controlling for autocorrelation and a control variable: $r_{t+1} = a + \beta \text{SSI}_t^{AR} + \gamma \text{CONT}_t + \epsilon_{t+1}$ in which r_{t+1} is the future index monthly return, SSI_t^{AR} is the Speculation Sentiment Index controlling for autocorrelation, β is the estimated coefficient on SSI_t^{AR}, CONT_t is a control variable, γ is the estimated coefficient on CONT_t, and ϵ_{t+1} is the error term. The control variables are index monthly return (r), cyclically adjusted earnings-to-price (CAEP), term spread (TERM), dividend-to-price (DP), short-rate (RATE), variance risk premium (VRP), intermediary capital risk factor (INTC), innovation to aggregate liquidity (Δ_LIQ), short interest (SHORT), VIX (VIX), Baker-Wurgler sentiment level (SENT), aligned investor sentiment level (HJTZ), closed-end fund discount (CEFD), consumer confidence level (CONF), change in consumer confidence (Δ_CONF), and investor lottery demand (FMAX). White standard errors are used to account for heteroscedasticity and t-statistics are reported, in parenthesis, below each estimated coefficient. The sample returns run from Dec. 2006 through Dec. 2019 (if the control variable is available through 2019). All variables, except for returns, are standardized.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
		r	CAEP	TERM	DP	RATE	VRP	INTC	Δ_LIQ	SHORT	VIX	SENT	HJTZ	CEFD	CONF	Δ_CONF	FMAX
Panel A: EW CRSP																	
SSI_t^{AR}	-1.69***	-1.76**	-1.67***	-1.70***	-1.68***	-1.68***	-1.22*	-1.59**	-1.55***	-1.58***	-1.78***	-1.59***	-1.61***	-1.71***	-1.70***	-1.69***	-1.51**
	(-2.87)	(-2.60)	(-2.75)	(-2.85)	(-2.74)	(-2.83)	(-1.91)	(-2.31)	(-2.81)	(-2.71)	(-2.94)	(-2.64)	(-2.77)	(-2.83)	(-2.84)	(-2.84)	(-2.40)
CONT_t		-0.02	0.89	0.22	0.95	-0.55*	1.23**	0.24	0.65	-0.90**	0.49	-1.01**	-0.37	0.91**	-0.17	0.19	0.46
		(-0.19)	(1.57)	(0.64)	(1.43)	(-1.79)	(2.58)	(0.42)	(1.24)	(-2.12)	(0.83)	(-2.57)	(-0.70)	(2.27)	(-0.34)	(0.45)	(0.90)
Adj. R^2	0.11	0.11	0.14	0.11	0.14	0.12	0.16	0.11	0.12	0.14	0.11	0.15	0.11	0.14	0.11	0.11	0.11
N	157	157	157	157	157	157	157	145	157	157	157	146	146	146	157	157	157
Panel B: VW CRSP																	
SSI_t^{AR}	-1.27**	-1.52**	-1.27**	-1.27**	-1.27**	-1.27**	-0.89	-1.27**	-1.11**	-1.21**	-1.28**	-1.22**	-1.14**	-1.28**	-1.26**	-1.27**	-1.27**
	(-2.44)	(-2.40)	(-2.39)	(-2.44)	(-2.40)	(-2.42)	(-1.53)	(-1.99)	(-2.44)	(-2.34)	(-2.43)	(-2.26)	(-2.31)	(-2.41)	(-2.43)	(-2.43)	(-2.18)
CONT_t		-0.10	0.29	-0.06	0.29	-0.29	0.99**	-0.00	0.74	-0.52	0.05	-0.50	-0.58	0.43	0.17	-0.04	0.02
		(-0.84)	(0.61)	(-0.18)	(0.52)	(-1.00)	(2.13)	(-0.00)	(1.40)	(-1.34)	(0.10)	(-1.43)	(-1.25)	(1.24)	(0.40)	(-0.10)	(0.04)
Adj. R^2	0.08	0.08	0.08	0.07	0.08	0.08	0.12	0.08	0.10	0.09	0.07	0.09	0.09	0.08	0.08	0.07	0.07
N	157	157	157	157	157	157	157	145	157	157	157	146	146	146	157	157	157
Panel C: S&P 500																	
SSI_t^{AR}	-1.12**	-1.30**	-1.11**	-1.11**	-1.11**	-1.11**	-0.76	-1.10*	-0.94**	-1.05**	-1.10**	-1.06**	-0.97**	-1.12**	-1.10**	-1.12**	-1.12**
	(-2.29)	(-2.18)	(-2.25)	(-2.29)	(-2.26)	(-2.27)	(-1.39)	(-1.82)	(-2.22)	(-2.19)	(-2.24)	(-2.11)	(-2.12)	(-2.25)	(-2.27)	(-2.29)	(-2.03)
CONT_t		-0.08	0.17	-0.08	0.17	-0.30	0.92**	0.03	0.78	-0.48	-0.08	-0.44	-0.62	0.41	0.24	-0.06	-0.00
		(-0.65)	(0.38)	(-0.26)	(0.32)	(-1.07)	(1.99)	(0.06)	(1.49)	(-1.29)	(-0.16)	(-1.32)	(-1.37)	(1.21)	(0.57)	(-0.16)	(-0.01)
Adj. R^2	0.06	0.06	0.06	0.06	0.06	0.06	0.10	0.06	0.08	0.07	0.06	0.07	0.08	0.07	0.06	0.06	0.06
N	157	157	157	157	157	157	157	145	157	157	157	146	146	146	157	157	157

Note: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

IA.7.2 SSI Orthogonal to Aggregate ETF Flows

ETF arbitrage activity (i.e., ETF flows) exhibits time trends across all funds. For example, since the mid 2000s, ETFs have exploded in popularity and the ETF industry as a whole has been characterized by ETF inflows. As a robustness test, I control for aggregate ETF flows in generating the time series of SSI. Specifically, I use,

$$\text{SSI}_t = a + \chi \text{ETFPCA1}_t + \text{SSI}_t^{flows}, \tag{IA90}$$

in which ETFPCA1_t is the first principal component that explains aggregate ETF flows. To form ETFPCA1_t, I take the largest 100 ETFs (based on June 2006 end-of-month market capitalizations) and form the first principal component that explains the joint variation in the covariance matrix of ETF share change (in which share change is measured as monthly percent change). The time series of SSI_t^{flows} forms the Speculation Sentiment Index orthogonal to aggregate ETF flows.

I repeat much of the regression analysis from Section IV and Section V but use SSI^{flows} in place of SSI. Table IA5 provides the results. Regression 1 presents the univariate regression results while Regressions 2-17 present the bivariate regression results with controls. The results in Table IA5 are qualitatively the same as the results in Table 3, Table 4, and Table 7. The analysis with SSI^{flows} provides evidence that the return predictability results are not due to the growth of ETF industry.

Table IA5: Return Predictability and SSI_t^{flow}

Regression 1 regresses the future CRSP equal weighted, CRSP value weighted, or S&P 500 index monthly returns on the Speculation Sentiment Index orthogonal to aggregate ETF flows: $r_{t+1} = a + \beta\text{SSI}_t^{flow} + \epsilon_{t+1}$ in which r_{t+1} is the future index monthly return, SSI_t^{flow} is the Speculation Sentiment Index orthogonal to aggregate ETF flows, β is the estimated coefficient on SSI_t^{flow}, and ϵ_{t+1} is the error term. Regressions 2-17 regress the future CRSP equal weighted, CRSP value weighted, or S&P 500 index monthly returns on the Speculation Sentiment Index orthogonal to aggregate ETF flows and a control variable: $r_{t+1} = a + \beta\text{SSI}_t^{flow} + \gamma\text{CONT}_t + \epsilon_{t+1}$ in which r_{t+1} is the future index monthly return, SSI_t^{flow} is the Speculation Sentiment Index orthogonal to aggregate ETF flows, β is the estimated coefficient on SSI_t^{flow}, CONT_t is a control variable, γ is the estimated coefficient on CONT_t, and ϵ_{t+1} is the error term. The control variables are index monthly return (r), cyclically adjusted earnings-to-price (CAEP), term spread (TERM), dividend-to-price (DP), short-rate (RATE), variance risk premium (VRP), intermediary capital risk factor (INTC), innovation to aggregate liquidity (Δ_LIQ), short interest (SHORT), VIX (VIX), Baker-Wurgler sentiment level (SENT), aligned investor sentiment level (HJTZ), closed-end fund discount (CEFD), consumer confidence level (CONF), change in consumer confidence (Δ_CONF), and investor lottery demand (FMAX). White standard errors are used to account for heteroscedasticity and t-statistics are reported, in parenthesis, below each estimated coefficient. The sample returns run from Nov. 2006 through Dec. 2019 (if the control variable is available through 2019). All variables, except for returns, are standardized.

	1	2 r	3 CAEP	4 TERM	5 DP	6 RATE	7 VRP	8 INTC	9 Δ_LIQ	10 SHORT	11 VIX	12 SENT	13 HJTZ	14 CEFD	15 CONF	16 Δ_CONF	17 FMAX
Panel A: EW CRSP																	
SSI_t^{flow}	-1.61***	-1.78**	-1.62***	-1.62***	-1.66***	-1.63***	-1.08*	-1.55**	-1.45***	-1.51***	-1.79***	-1.54***	-1.52***	-1.69***	-1.63***	-1.60***	-1.42**
	(-2.93)	(-2.61)	(-2.87)	(-2.93)	(-2.92)	(-2.97)	(-1.77)	(-2.22)	(-2.88)	(-2.79)	(-3.16)	(-2.76)	(-2.83)	(-3.04)	(-2.94)	(-2.86)	(-2.30)
CONT_t		-0.05	0.95	0.23	1.04	-0.57*	1.23**	0.13	0.66	-0.90**	0.67	-1.02***	-0.29	0.97**	-0.23	0.11	0.39
		(-0.39)	(1.65)	(0.69)	(1.57)	(-1.96)	(2.40)	(0.21)	(1.20)	(-2.18)	(1.11)	(-2.65)	(-0.55)	(2.45)	(-0.44)	(0.27)	(0.70)
Adj. R^2	0.10	0.10	0.13	0.10	0.14	0.11	0.15	0.10	0.11	0.13	0.11	0.14	0.10	0.13	0.10	0.09	0.10
N	158	158	158	158	158	158	158	146	158	158	158	147	147	147	158	158	158
Panel B: VW CRSP																	
SSI_t^{flow}	-1.20**	-1.56**	-1.20**	-1.20**	-1.22**	-1.21**	-0.77	-1.24*	-1.02**	-1.14**	-1.24**	-1.16**	-1.04**	-1.23**	-1.18**	-1.20**	-1.23**
	(-2.48)	(-2.50)	(-2.46)	(-2.47)	(-2.49)	(-2.50)	(-1.38)	(-1.93)	(-2.46)	(-2.39)	(-2.49)	(-2.34)	(-2.26)	(-2.53)	(-2.45)	(-2.44)	(-2.17)
CONT_t		-0.13	0.33	-0.04	0.36	-0.32	1.00**	-0.09	0.74	-0.52	0.16	-0.51	-0.53	0.48	0.13	-0.09	-0.06
		(-1.02)	(0.70)	(-0.13)	(0.65)	(-1.16)	(2.03)	(-0.16)	(1.38)	(-1.40)	(0.33)	(-1.51)	(-1.13)	(1.41)	(0.30)	(-0.25)	(-0.12)
Adj. R^2	0.07	0.08	0.07	0.07	0.07	0.07	0.11	0.07	0.09	0.08	0.07	0.08	0.08	0.08	0.07	0.07	0.07
N	158	158	158	158	158	158	158	146	158	158	158	147	147	147	158	158	158
Panel C: S&P 500																	
SSI_t^{flow}	-1.06**	-1.33**	-1.06**	-1.06**	-1.07**	-1.08**	-0.67	-1.08*	-0.87**	-1.01**	-1.07**	-1.02**	-0.89**	-1.09**	-1.04**	-1.07**	-1.10**
	(-2.36)	(-2.28)	(-2.35)	(-2.35)	(-2.38)	(-2.39)	(-1.26)	(-1.79)	(-2.26)	(-2.28)	(-2.29)	(-2.23)	(-2.07)	(-2.42)	(-2.31)	(-2.33)	(-2.05)
CONT_t		-0.10	0.21	-0.06	0.24	-0.34	0.92*	-0.05	0.78	-0.49	0.02	-0.46	-0.57	0.46	0.20	-0.11	-0.08
		(-0.80)	(0.46)	(-0.19)	(0.44)	(-1.25)	(1.88)	(-0.10)	(1.47)	(-1.34)	(0.05)	(-1.41)	(-1.25)	(1.39)	(0.48)	(-0.31)	(-0.17)
Adj. R^2	0.06	0.06	0.06	0.05	0.06	0.06	0.09	0.06	0.08	0.07	0.05	0.07	0.07	0.07	0.05	0.05	0.05
N	158	158	158	158	158	158	158	146	158	158	158	147	147	147	158	158	158

Note: $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$

IA.7.3 SSI Orthogonal to Aggregate Macro Conditions

ETF arbitrage activity (i.e., ETF flows) is an equilibrium outcome and reflects, among other market conditions, the cost of arbitrage capital. As a robustness test, I control for several macro variables in generating the time series of SSI. Specifically, I use,

$$\text{SSI}_t = a + \chi\text{CONTROLS}_t + \text{SSI}_t^{\perp}, \tag{IA91}$$

in which controls$_t$ consists of short interest (SHORT), VIX (VIX), and the intermediary capital risk factor (INTC). $\text{SSI}_t^{\perp}$ forms the Speculation Sentiment Index orthogonal to macro conditions.

I repeat much of the regression analysis from Section IV and Section V but use $\text{SSI}^{\perp}$ in place of SSI. Table IA6 provides the results. Regression 1 presents the univariate regression results while Regressions 2-17 present the bivariate regression results with controls. The results in Table IA6 are qualitatively the same as the results in Table 3, Table 4, and Table 7. The analysis with $\text{SSI}^{\perp}$ provides evidence that the return predictability results are not due to variation in the costs of arbitrage capital.

IA.7.4 Dollar Flow SSI

The main specification of SSI is based on percent changes in shares outstanding for the leveraged ETFs, rather than dollar changes. Using percent changes in shares outstanding has the attractive feature that they are likely more stationary than dollar changes. However, in this robustness test, I compute a dollar flow measure. Specifically, I

Table IA6: Return Predictability and $\text{SSI}_t^{\perp}$

Regression 1 regresses the future CRSP equal weighted, CRSP value weighted, or S&P 500 index monthly returns on the Speculation Sentiment Index orthogonal to macro conditions: $r_{t+1} = a + \beta\text{SSI}_t^{\perp} + \epsilon_{t+1}$ in which r_{t+1} is the future index monthly return, $\text{SSI}_t^{\perp}$ is the Speculation Sentiment Index orthogonal to macro conditions, β is the estimated coefficient on $\text{SSI}_t^{\perp}$, and ϵ_{t+1} is the error term. Regressions 2-17 regress the future CRSP equal weighted, CRSP value weighted, or S&P 500 index monthly returns on the Speculation Sentiment Index orthogonal to macro conditions and a control variable: $r_{t+1} = a + \beta\text{SSI}_t^{\perp} + \gamma\text{CONT}_t + \epsilon_{t+1}$ in which r_{t+1} is the future index monthly return, $\text{SSI}_t^{\perp}$ is the Speculation Sentiment Index orthogonal to macro conditions, β is the estimated coefficient on $\text{SSI}_t^{\perp}$, CONT_t is a control variable, γ is the estimated coefficient on CONT_t, and ϵ_{t+1} is the error term. The control variables are index monthly return (r), cyclically adjusted earnings-to-price (CAEP), term spread (TERM), dividend-to-price (DP), short-rate (RATE), variance risk premium (VRP), intermediary capital risk factor (INTC), innovation to aggregate liquidity (Δ_LIQ), short interest (SHORT), VIX (VIX), Baker-Wurgler sentiment level (SENT), aligned investor sentiment level (HJTZ), closed-end fund discount (CEFD), consumer confidence level (CONF), change in consumer confidence (Δ_CONF), and investor lottery demand (FMAX). White standard errors are used to account for heteroscedasticity and t-statistics are reported, in parenthesis, below each estimated coefficient. The sample returns run from Nov. 2006 through Dec. 2018 (note, 2019 data is not included as $\text{SSI}^{\perp}$ is obtained using the intermediate capital risk factor data which is available only through Nov. 2018). All variables, except for returns, are standardized.

	1	2 r	3 CAEP	4 TERM	5 DP	6 RATE	7 VRP	8 INTC	9 Δ_LIQ	10 SHORT	11 VIX	12 SENT	13 HJTZ	14 CEFD	15 CONF	16 Δ_CONF	17 FMAX
Panel A: EW CRSP																	
$\text{SSI}_t^{\perp}$	-1.41**	-1.14*	-1.34*	-1.47**	-1.36**	-1.54**	-0.84	-1.41**	-1.21**	-1.41**	-1.41**	-1.44**	-1.32**	-1.60**	-1.41**	-1.42**	-1.22**
	(-2.18)	(-1.86)	(-1.95)	(-2.25)	(-1.99)	(-2.41)	(-1.25)	(-2.26)	(-2.16)	(-2.36)	(-2.16)	(-2.28)	(-2.15)	(-2.50)	(-2.16)	(-2.21)	(-1.98)
CONT_t		0.17	0.96	0.51	0.91	-0.85***	1.54***	0.94*	1.09*	-1.19***	0.11	-1.12***	-0.46	1.05***	-0.15	0.31	0.88*
		(1.50)	(1.52)	(1.52)	(1.22)	(-2.72)	(3.28)	(1.94)	(1.93)	(-2.72)	(0.17)	(-2.94)	(-0.87)	(2.67)	(-0.28)	(0.67)	(1.66)
Adj. R^2	0.07	0.09	0.10	0.08	0.10	0.10	0.15	0.10	0.10	0.13	0.07	0.12	0.08	0.11	0.07	0.07	0.10
N	146	146	146	146	146	146	146	146	146	146	146	146	146	146	146	146	146
Panel B: VW CRSP																	
$\text{SSI}_t^{\perp}$	-1.10*	-1.00*	-1.08*	-1.13*	-1.09*	-1.19**	-0.66	-1.10*	-0.91*	-1.10**	-1.10*	-1.12*	-0.98*	-1.20**	-1.10*	-1.10*	-1.02*
	(-1.89)	(-1.71)	(-1.79)	(-1.91)	(-1.83)	(-2.03)	(-1.05)	(-1.91)	(-1.83)	(-2.00)	(-1.91)	(-1.93)	(-1.80)	(-2.05)	(-1.91)	(-1.89)	(-1.75)
CONT_t		0.08	0.36	0.24	0.26	-0.58*	1.22**	0.55	1.08*	-0.78*	-0.22	-0.59*	-0.63	0.55	0.14	0.00	0.37
		(0.77)	(0.71)	(0.77)	(0.43)	(-1.91)	(2.58)	(1.51)	(1.89)	(-1.95)	(-0.43)	(-1.79)	(-1.34)	(1.59)	(0.33)	(0.00)	(0.80)
Adj. R^2	0.06	0.06	0.06	0.05	0.05	0.07	0.12	0.07	0.10	0.08	0.05	0.07	0.07	0.07	0.05	0.05	0.06
N	146	146	146	146	146	146	146	146	146	146	146	146	146	146	146	146	146
Panel C: S&P 500																	
$\text{SSI}_t^{\perp}$	-0.95*	-0.85	-0.93	-0.97*	-0.94*	-1.03*	-0.53	-0.95*	-0.75	-0.95*	-0.95*	-0.96*	-0.82	-1.04*	-0.95*	-0.94*	-0.88
	(-1.71)	(-1.53)	(-1.64)	(-1.73)	(-1.67)	(-1.86)	(-0.89)	(-1.72)	(-1.57)	(-1.81)	(-1.74)	(-1.75)	(-1.58)	(-1.87)	(-1.73)	(-1.71)	(-1.56)
CONT_t		0.09	0.26	0.23	0.15	-0.58**	1.13**	0.51	1.09*	-0.73*	-0.31	-0.52*	-0.66	0.51	0.19	-0.04	0.31
		(0.81)	(0.52)	(0.73)	(0.26)	(-1.99)	(2.41)	(1.47)	(1.94)	(-1.91)	(-0.63)	(-1.66)	(-1.44)	(1.53)	(0.45)	(-0.12)	(0.68)
Adj. R^2	0.04	0.04	0.04	0.04	0.04	0.06	0.10	0.05	0.09	0.07	0.04	0.05	0.06	0.05	0.04	0.04	0.04
N	146	146	146	146	146	146	146	146	146	146	146	146	146	146	146	146	146

Note: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

use,

$$\text{(IA92)}\quad \mathrm{SSI}^{\$}_t = \sum_{i\in J}(\mathrm{SO}_{i,t} - \mathrm{SO}_{i,t-1})\left(\frac{P_{i,t}+P_{i,t-1}}{2}\right) - \sum_{i\in K}(\mathrm{SO}_{i,t} - \mathrm{SO}_{i,t-1})\left(\frac{P_{i,t}+P_{i,t-1}}{2}\right).$$

Equation (IA92) represents the difference in dollar flows into the leveraged-long ETFs and leveraged-short ETFs, in which changes in shares outstanding are weighted by the average price.

I repeat much of the regression analysis from Section IV and Section V but use $\mathrm{SSI}^{\$}$ in place of SSI. Table IA7 provides the results. Regression 1 presents the univariate regression results while Regressions 2-17 present the bivariate regression results with controls. The results in Table IA7 are qualitatively the same as the results in Table 3, Table 4, and Table 7. The analysis with $\mathrm{SSI}^{\$}$ provides evidence that the return predictability results are robust to measuring Speculation Sentiment with dollar flows.

IA.7.5 Evolving SSI

The baseline specification of SSI is restricted to the original set of leveraged ETFs. Since the introduction of the ProShares funds in 2006, there have been many -3x, -2x, 2x, and 3x leveraged ETFs launched. As a robustness test, I form an evolving version of SSI. Specifically, I include any leveraged ETF pair that follows either the Dow Jones Industrial Average, NASDAQ-100 index, and the S&P 500 index. In total, there are 7 ETF pairs (14 funds in total). Each month, a leveraged-long, index-level ETF share change is computed

Table IA7: Return Predictability and $\text{SSI}_t^{\$}$

Regression 1 regresses the future CRSP equal weighted, CRSP value weighted, or S&P 500 index monthly returns on the dollar flow Speculation Sentiment Index: $r_{t+1} = a + \beta \text{SSI}_t^{\$} + \epsilon_{t+1}$ in which r_{t+1} is the future index monthly return, $\text{SSI}_t^{\$}$ is the dollar flow Speculation Sentiment Index, β is the estimated coefficient on $\text{SSI}_t^{\$}$, and ϵ_{t+1} is the error term. Regressions 2-17 regress the future CRSP equal weighted, CRSP value weighted, or S&P 500 index monthly returns on the dollar flow Speculation Sentiment Index and a control variable: $r_{t+1} = a + \beta \text{SSI}_t^{\$} + \gamma \text{CONT}_t + \epsilon_{t+1}$ in which r_{t+1} is the future index monthly return, $\text{SSI}_t^{\$}$ is the dollar flow Speculation Sentiment Index, β is the estimated coefficient on $\text{SSI}_t^{\$}$, CONT_t is a control variable, γ is the estimated coefficient on CONT_t, and ϵ_{t+1} is the error term. The control variables are index monthly return (r), cyclically adjusted earnings-to-price (CAEP), term spread (TERM), dividend-to-price (DP), short-rate (RATE), variance risk premium (VRP), intermediary capital risk factor (INTC), innovation to aggregate liquidity (Δ_LIQ), short interest (SHORT), VIX (VIX), Baker-Wurgler sentiment level (SENT), aligned investor sentiment level (HJTZ), closed-end fund discount (CEFD), consumer confidence level (CONF), change in consumer confidence (Δ_CONF), and investor lottery demand (FMAX). White standard errors are used to account for heteroscedasticity and t-statistics are reported, in parenthesis, below each estimated coefficient. The sample returns run from Nov. 2006 through Dec. 2019 (if the control variable is available through 2019). All variables, except for returns, are standardized.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
		r	CAEP	TERM	DP	RATE	VRP	INTC	Δ_LIQ	SHORT	VIX	SENT	HJTZ	CEFD	CONF	Δ_CONF	FMAX
Panel A: EW CRSP																	
$\text{SSI}_t^{\$}$	-1.56***	-1.64**	-1.54**	-1.55***	-1.59***	-1.55***	-1.03	-1.47**	-1.42***	-1.46**	-1.69***	-1.44**	-1.45**	-1.60***	-1.57***	-1.55***	-1.36**
	(-2.70)	(-2.44)	(-2.60)	(-2.68)	(-2.69)	(-2.70)	(-1.63)	(-2.03)	(-2.73)	(-2.59)	(-2.90)	(-2.46)	(-2.60)	(-2.79)	(-2.71)	(-2.64)	(-2.03)
CONT_t		-0.03	0.89	0.12	0.99	-0.49	1.26**	0.16	0.78	-0.90**	0.56	-0.97**	-0.33	0.92**	-0.16	0.13	0.38
		(-0.20)	(1.51)	(0.36)	(1.48)	(-1.65)	(2.41)	(0.25)	(1.42)	(-2.19)	(0.93)	(-2.44)	(-0.61)	(2.29)	(-0.31)	(0.32)	(0.66)
Adj. R^2	0.09	0.09	0.12	0.09	0.13	0.10	0.14	0.09	0.11	0.12	0.10	0.12	0.09	0.12	0.09	0.09	0.09
N	158	158	158	158	158	158	158	146	158	158	158	147	147	147	158	158	158
Panel B: VW CRSP																	
$\text{SSI}_t^{\$}$	-1.13**	-1.36**	-1.12**	-1.13**	-1.14**	-1.13**	-0.69	-1.13*	-0.98**	-1.07**	-1.15**	-1.06**	-0.95**	-1.13**	-1.11**	-1.13**	-1.15*
	(-2.23)	(-2.24)	(-2.18)	(-2.23)	(-2.24)	(-2.22)	(-1.21)	(-1.70)	(-2.31)	(-2.16)	(-2.26)	(-2.04)	(-2.00)	(-2.25)	(-2.22)	(-2.20)	(-1.89)
CONT_t		-0.09	0.29	-0.12	0.32	-0.26	1.04**	-0.04	0.83	-0.53	0.09	-0.48	-0.57	0.44	0.18	-0.07	-0.05
		(-0.73)	(0.60)	(-0.39)	(0.59)	(-0.94)	(2.10)	(-0.08)	(1.57)	(-1.41)	(0.17)	(-1.37)	(-1.21)	(1.29)	(0.42)	(-0.20)	(-0.10)
Adj. R^2	0.06	0.06	0.06	0.06	0.06	0.06	0.10	0.06	0.08	0.07	0.06	0.07	0.07	0.07	0.06	0.06	0.06
N	158	158	158	158	158	158	158	146	158	158	158	147	147	147	158	158	158
Panel C: S&P 500																	
$\text{SSI}_t^{\$}$	-0.98**	-1.13**	-0.98**	-0.98**	-0.99**	-0.98**	-0.58	-0.96	-0.82**	-0.92**	-0.97**	-0.91*	-0.79*	-0.99**	-0.96**	-0.99**	-1.01*
	(-2.08)	(-2.01)	(-2.04)	(-2.08)	(-2.09)	(-2.07)	(-1.07)	(-1.53)	(-2.10)	(-2.00)	(-2.05)	(-1.89)	(-1.79)	(-2.10)	(-2.06)	(-2.06)	(-1.74)
CONT_t		-0.06	0.18	-0.13	0.21	-0.28	0.97*	-0.00	0.86	-0.49	-0.05	-0.43	-0.61	0.42	0.24	-0.09	-0.06
		(-0.52)	(0.37)	(-0.43)	(0.38)	(-1.06)	(1.98)	(-0.00)	(1.65)	(-1.36)	(-0.10)	(-1.28)	(-1.34)	(1.28)	(0.59)	(-0.27)	(-0.12)
Adj. R^2	0.05	0.05	0.04	0.04	0.05	0.05	0.09	0.04	0.07	0.06	0.04	0.05	0.06	0.05	0.05	0.04	0.04
N	158	158	158	158	158	158	158	146	158	158	158	147	147	147	158	158	158

Note: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

by taking a weighted average of each leveraged-long ETF's share change (in which weights are determined by monthly ETF market capitalizations) for each of the three indices (i.e, the Dow Jones Industrial Average, NASDAQ-100 index, and the S&P 500 index). Similarly, a leveraged-short, index-level ETF share change is computed by taking a weighted average of each leveraged-short ETF's share change. Then, as in equation (2), the net change is computed by taking the difference between the leveraged-long and the leveraged-short index changes (forming SSI*). The evolving SSI* allows for the index to reflect the introduction of new leveraged ETFs. Furthermore, by weighting share change within benchmark index category by market capitalization, investor preferences are also reflected in the evolving SSI (i.e., the more popular and larger ETFs exhibit greater representation in the index).

I repeat much of the regression analysis from Section IV and Section V but use SSI* in place of SSI. Table IA8 provides the results. Regression 1 presents the univariate regression results while Regressions 2-17 present the bivariate regression results with controls. The results in Table IA8 are qualitatively the same as the results in Table 3, Table 4, and Table 7. The results show that accounting for new leveraged ETFs provides similar insights as to using the original six ProShares ETFs.

IA.7.6 Long Component and Short Component Separated

SSI is constructed by taking the difference between leveraged-long ETFs' share change and leveraged-short ETFs' share change, as seen in equation (2). The theoretical

Table IA8: Return Predictability and SSI_t^*

Regression 1 regresses the future CRSP equal weighted, CRSP value weighted, or S&P 500 index monthly returns on the evolving Speculation Sentiment Index: $r_{t+1} = a + \beta \text{SSI}_t^* + \epsilon_{t+1}$ in which r_{t+1} is the future index monthly return, SSI_t^* is the evolving Speculation Sentiment Index, β is the estimated coefficient on SSI_t^*, and ϵ_{t+1} is the error term. Regressions 2-17 regress the future CRSP equal weighted, CRSP value weighted, or S&P 500 index monthly returns on the evolving Speculation Sentiment Index and a control variable: $r_{t+1} = a + \beta \text{SSI}_t^* + \gamma \text{CONT}_t + \epsilon_{t+1}$ in which r_{t+1} is the future index monthly return, SSI_t^* is the evolving Speculation Sentiment Index, β is the estimated coefficient on SSI_t^*, CONT_t is a control variable, γ is the estimated coefficient on CONT_t, and ϵ_{t+1} is the error term. The control variables are index monthly return (r), cyclically adjusted earnings-to-price (CAEP), term spread (TERM), dividend-to-price (DP), short-rate (RATE), variance risk premium (VRP), intermediary capital risk factor (INTC), innovation to aggregate liquidity (Δ_LIQ), short interest (SHORT), VIX (VIX), Baker-Wurgler sentiment level (SENT), aligned investor sentiment level (HJTZ), closed-end fund discount (CEFD), consumer confidence level (CONF), change in consumer confidence (Δ_CONF), and investor lottery demand (FMAX). White standard errors are used to account for heteroscedasticity and t-statistics are reported, in parenthesis, below each estimated coefficient. The sample returns run from Nov. 2006 through Dec. 2019 (if the control variable is available through 2019). All variables, except for returns, are standardized.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
		r	CAEP	TERM	DP	RATE	VRP	INTC	Δ_LIQ	SHORT	VIX	SENT	HJTZ	CEFD	CONF	Δ_CONF	FMAX
Panel A: EW CRSP																	
SSI_t^*	-1.63***	-1.94***	-1.66***	-1.64***	-1.69***	-1.64***	-1.12*	-1.70**	-1.47***	-1.53***	-1.84***	-1.57***	-1.58***	-1.72***	-1.66***	-1.62***	-1.46**
	(-3.09)	(-2.85)	(-3.06)	(-3.08)	(-3.08)	(-3.08)	(-1.92)	(-2.44)	(-3.06)	(-2.94)	(-3.34)	(-2.88)	(-2.99)	(-3.14)	(-3.08)	(-3.02)	(-2.43)
CONT_t		-0.09	0.97*	0.24	1.04	-0.52*	1.22**	0.00	0.64	-0.90**	0.72	-0.99**	-0.27	0.92**	-0.24	0.13	0.35
		(-0.64)	(1.71)	(0.70)	(1.59)	(-1.77)	(2.47)	(0.00)	(1.14)	(-2.20)	(1.19)	(-2.55)	(-0.51)	(2.34)	(-0.46)	(0.31)	(0.61)
Adj. R^2	0.10	0.10	0.14	0.10	0.14	0.11	0.15	0.11	0.11	0.13	0.12	0.14	0.10	0.13	0.10	0.10	0.10
N	158	158	158	158	158	158	158	146	158	158	158	147	147	147	158	158	158
Panel B: VW CRSP																	
SSI_t^*	-1.22***	-1.74***	-1.23***	-1.22***	-1.24***	-1.23***	-0.80	-1.36**	-1.04***	-1.17**	-1.28***	-1.19**	-1.08**	-1.26***	-1.21**	-1.23**	-1.27**
	(-2.63)	(-2.67)	(-2.62)	(-2.62)	(-2.65)	(-2.62)	(-1.51)	(-2.10)	(-2.66)	(-2.55)	(-2.66)	(-2.44)	(-2.38)	(-2.62)	(-2.59)	(-2.59)	(-2.32)
CONT_t		-0.18	0.35	-0.03	0.36	-0.28	0.99**	-0.19	0.73	-0.52	0.20	-0.49	-0.52	0.45	0.12	-0.08	-0.10
		(-1.25)	(0.75)	(-0.11)	(0.67)	(-1.02)	(2.06)	(-0.34)	(1.34)	(-1.40)	(0.40)	(-1.43)	(-1.10)	(1.31)	(0.28)	(-0.22)	(-0.20)
Adj. R^2	0.07	0.09	0.07	0.07	0.08	0.07	0.11	0.08	0.09	0.08	0.07	0.08	0.08	0.08	0.07	0.07	0.07
N	158	158	158	158	158	158	158	146	158	158	158	147	147	147	158	158	158
Panel C: S&P 500																	
SSI_t^*	-1.08**	-1.49**	-1.09**	-1.08**	-1.10**	-1.09**	-0.70	-1.18*	-0.90**	-1.03**	-1.10**	-1.04**	-0.92**	-1.11**	-1.06**	-1.09**	-1.15**
	(-2.51)	(-2.40)	(-2.51)	(-2.50)	(-2.53)	(-2.50)	(-1.39)	(-1.92)	(-2.45)	(-2.43)	(-2.44)	(-2.31)	(-2.18)	(-2.49)	(-2.46)	(-2.48)	(-2.18)
CONT_t		-0.15	0.23	-0.05	0.24	-0.30	0.91*	-0.14	0.76	-0.49	0.05	-0.44	-0.56	0.43	0.19	-0.10	-0.12
		(-1.00)	(0.51)	(-0.18)	(0.45)	(-1.12)	(1.90)	(-0.26)	(1.44)	(-1.35)	(0.11)	(-1.34)	(-1.22)	(1.31)	(0.46)	(-0.29)	(-0.24)
Adj. R^2	0.06	0.07	0.06	0.06	0.06	0.06	0.10	0.06	0.08	0.07	0.06	0.07	0.07	0.07	0.06	0.06	0.06
N	158	158	158	158	158	158	158	146	158	158	158	147	147	147	158	158	158

Note: $^{*}\ p < 0.1$; $^{**}\ p < 0.05$; $^{***}\ p < 0.01$

underpinning for the index's construction is that it captures the net bullish-bearish speculation sentiment, that is, only when there is consensus among speculators is the index significantly bullish or bearish. The netting in equation (2) does not allow one to examine the predictability coming from *only* leveraged-long ETFs, nor does it allow one to examine the predictability coming from *only* leveraged-short ETFs. It is natural to consider each separately. Moreover, the model in Section IA.2 of this Internet Appendix suggests that leveraged-long ETF share changes should *negatively* predict returns and leveraged-short ETF share changes should *positively* predict returns (see Remark IA1). Define SSI^L as the long-component of SSI and define SSI^S as the short-component.

I repeat much of the regression analysis from Section IV and Section V but use SSI^L and SSI^S in place of SSI. Table IA9 provides the results using SSI^L. Regression 1 presents the univariate regression results while Regressions 2-17 present the bivariate regression results with controls. The results in Table IA9 are qualitatively the same as the results in Table 3, Table 4, and Table 7, albeit slightly weaker. Table IA10 provides the results using SSI^S. The results in Table IA9 demonstrate *positive* return predictability (consistent with Remark IA1 from Section IA.2), but are weaker than those using SSI or SSI^L. Collectively, Table IA9 and Table IA10 show that both SSI^L and SSI^S provide predictability, but both are weaker predictors than the main specification SSI. The weaker predictability is consistent with Remark IA3 from Section IA.2, that is, netting the leveraged-long and leveraged-short share changes attenuates noise and provides a cleaner measurement of speculative demand shocks.

Table IA9: Return Predictability and SSI_t^L

Regression 1 regresses the future CRSP equal weighted, CRSP value weighted, or S&P 500 index monthly returns on the leveraged-long Speculation Sentiment Index: $r_{t+1} = a + \beta \mathrm{SSI}_t^L + \epsilon_{t+1}$ in which r_{t+1} is the future index monthly return, SSI_t^L is the leveraged-long Speculation Sentiment Index, β is the estimated coefficient on SSI_t^L, and ϵ_{t+1} is the error term. Regressions 2-17 regress the future CRSP equal weighted, CRSP value weighted, or S&P 500 index monthly returns on the leveraged-long Speculation Sentiment Index and a control variable: $r_{t+1} = a + \beta \mathrm{SSI}_t^L + \gamma \mathrm{CONT}_t + \epsilon_{t+1}$ in which r_{t+1} is the future index monthly return, SSI_t^L is the leveraged-long Speculation Sentiment Index, β is the estimated coefficient on SSI_t^L, CONT_t is a control variable, γ is the estimated coefficient on CONT_t, and ϵ_{t+1} is the error term. The control variables are index monthly return (r), cyclically adjusted earnings-to-price (CAEP), term spread (TERM), dividend-to-price (DP), short-rate (RATE), variance risk premium (VRP), intermediary capital risk factor (INTC), innovation to aggregate liquidity (Δ_LIQ), short interest (SHORT), VIX (VIX), Baker-Wurgler sentiment level (SENT), aligned investor sentiment level (HJTZ), closed-end fund discount (CEFD), consumer confidence level (CONF), change in consumer confidence (Δ_CONF), and investor lottery demand (FMAX). White standard errors are used to account for heteroscedasticity and t-statistics are reported, in parenthesis, below each estimated coefficient. The sample returns run from Nov. 2006 through Dec. 2019 (if the control variable is available through 2019). All variables, except for returns, are standardized.

Panel A: EW CRSP

	1	2 r	3 CAEP	4 TERM	5 DP	6 RATE	7 VRP	8 INTC	9 Δ_LIQ	10 SHORT	11 VIX	12 SENT	13 HJTZ	14 CEFD	15 CONF	16 Δ_CONF	17 FMAX
SSI_t^L	-1.64***	-1.63***	-1.68***	-1.64***	-1.73***	-1.61***	-1.02	-1.51**	-1.48***	-1.45**	-1.87***	-1.44**	-1.55***	-1.55***	-1.67***	-1.63***	-1.44***
	(-3.18)	(-2.79)	(-3.07)	(-3.14)	(-3.09)	(-2.92)	(-1.44)	(-2.39)	(-3.22)	(-2.60)	(-3.21)	(-2.45)	(-3.16)	(-2.71)	(-3.14)	(-3.06)	(-2.76)
CONT_t		0.01	0.99*	-0.03	1.09*	-0.11	1.13*	0.26	0.64	-0.64	0.75	-0.73*	-0.39	0.57	-0.26	0.08	0.56
		(0.05)	(1.78)	(-0.08)	(1.71)	(-0.32)	(1.90)	(0.44)	(1.18)	(-1.49)	(1.26)	(-1.66)	(-0.78)	(1.27)	(-0.52)	(0.20)	(1.10)
Adj. R^2	0.10	0.10	0.14	0.10	0.15	0.10	0.14	0.10	0.11	0.11	0.12	0.12	0.11	0.11	0.10	0.10	0.11
N	158	158	158	158	158	158	158	146	158	158	158	147	147	147	158	158	158

Panel B: VW CRSP

	1	2 r	3 CAEP	4 TERM	5 DP	6 RATE	7 VRP	8 INTC	9 Δ_LIQ	10 SHORT	11 VIX	12 SENT	13 HJTZ	14 CEFD	15 CONF	16 Δ_CONF	17 FMAX
SSI_t^L	-1.23***	-1.41**	-1.25***	-1.26***	-1.26***	-1.24**	-0.72	-1.19**	-1.05***	-1.14**	-1.30**	-1.14**	-1.08**	-1.18**	-1.22**	-1.24**	-1.19**
	(-2.66)	(-2.57)	(-2.64)	(-2.70)	(-2.66)	(-2.53)	(-1.09)	(-2.03)	(-2.77)	(-2.33)	(-2.58)	(-2.20)	(-2.53)	(-2.38)	(-2.60)	(-2.60)	(-2.47)
CONT_t		-0.08	0.36	-0.23	0.39	0.03	0.94	0.02	0.73	-0.31	0.22	-0.28	-0.60	0.18	0.10	-0.11	0.11
		(-0.69)	(0.79)	(-0.74)	(0.74)	(0.10)	(1.64)	(0.05)	(1.37)	(-0.81)	(0.43)	(-0.73)	(-1.31)	(0.47)	(0.24)	(-0.31)	(0.24)
Adj. R^2	0.08	0.07	0.08	0.07	0.08	0.07	0.10	0.07	0.09	0.07	0.07	0.07	0.09	0.07	0.07	0.07	0.07
N	158	158	158	158	158	158	158	146	158	158	158	147	147	147	158	158	158

Panel C: S&P 500

	1	2 r	3 CAEP	4 TERM	5 DP	6 RATE	7 VRP	8 INTC	9 Δ_LIQ	10 SHORT	11 VIX	12 SENT	13 HJTZ	14 CEFD	15 CONF	16 Δ_CONF	17 FMAX
SSI_t^L	-1.12***	-1.28**	-1.13***	-1.15***	-1.15***	-1.12**	-0.66	-1.08**	-0.94***	-1.04**	-1.15**	-1.03**	-0.96**	-1.07**	-1.10**	-1.14***	-1.10**
	(-2.66)	(-2.50)	(-2.65)	(-2.72)	(-2.67)	(-2.49)	(-1.07)	(-1.98)	(-2.72)	(-2.32)	(-2.49)	(-2.19)	(-2.47)	(-2.35)	(-2.58)	(-2.62)	(-2.46)
CONT_t		-0.07	0.24	-0.23	0.27	-0.02	0.85	0.03	0.75	-0.29	0.08	-0.25	-0.62	0.19	0.17	-0.13	0.05
		(-0.61)	(0.54)	(-0.76)	(0.52)	(-0.06)	(1.49)	(0.06)	(1.45)	(-0.78)	(0.16)	(-0.67)	(-1.40)	(0.51)	(0.42)	(-0.38)	(0.12)
Adj. R^2	0.07	0.06	0.06	0.06	0.06	0.06	0.09	0.06	0.08	0.06	0.06	0.06	0.08	0.06	0.06	0.06	0.06
N	158	158	158	158	158	158	158	146	158	158	158	147	147	147	158	158	158

Note: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Table IA10: Return Predictability and SSI_t^S

Regression 1 regresses the future CRSP equal weighted, CRSP value weighted, or S&P 500 index monthly returns on the leveraged-short Speculation Sentiment Index: $r_{t+1} = a + \beta \mathrm{SSI}_t^S + \epsilon_{t+1}$ in which r_{t+1} is the future index monthly return, SSI_t^S is the leveraged-short Speculation Sentiment Index, β is the estimated coefficient on SSI_t^S, and ϵ_{t+1} is the error term. Regressions 2-17 regress the future CRSP equal weighted, CRSP value weighted, or S&P 500 index monthly returns on the leveraged-short Speculation Sentiment Index and a control variable: $r_{t+1} = a + \beta \mathrm{SSI}_t^S + \gamma \mathrm{CONT}_t + \epsilon_{t+1}$ in which r_{t+1} is the future index monthly return, SSI_t^S is the leveraged-short Speculation Sentiment Index, β is the estimated coefficient on SSI_t^S, CONT_t is a control variable, γ is the estimated coefficient on CONT_t, and ϵ_{t+1} is the error term. The control variables are index monthly return (r), cyclically adjusted earnings-to-price (CAEP), term spread (TERM), dividend-to-price (DP), short-rate (RATE), variance risk premium (VRP), intermediary capital risk factor (INTC), innovation to aggregate liquidity (Δ_LIQ), short interest (SHORT), VIX (VIX), Baker-Wurgler sentiment level (SENT), aligned investor sentiment level (HJTZ), closed-end fund discount (CEFD), consumer confidence level (CONF), change in consumer confidence (Δ_CONF), and investor lottery demand (FMAX). White standard errors are used to account for heteroscedasticity and t-statistics are reported, in parenthesis, below each estimated coefficient. The sample returns run from Nov. 2006 through Dec. 2019 (if the control variable is available through 2019). All variables, except for returns, are standardized.

	1	2 r	3 CAEP	4 TERM	5 DP	6 RATE	7 VRP	8 INTC	9 Δ_LIQ	10 SHORT	11 VIX	12 SENT	13 HJTZ	14 CEFD	15 CONF	16 Δ_CONF	17 FMAX
Panel A: EW CRSP																	
SSI_t^S	0.97**	0.67	0.95**	1.04**	0.97**	1.27***	0.77*	0.80	0.85*	1.13***	1.01**	1.19***	0.85*	1.37***	0.97**	0.97**	0.64
	(2.13)	(1.54)	(2.24)	(2.26)	(2.33)	(2.72)	(1.94)	(1.65)	(1.96)	(2.76)	(2.29)	(2.89)	(1.86)	(3.21)	(2.15)	(2.11)	(1.47)
CONT_t		0.12	0.90	0.35	0.94	-0.91***	1.59***	0.64	0.96	-1.22***	0.29	-1.33***	-0.53	1.26***	-0.09	0.23	0.82
		(0.90)	(1.45)	(0.97)	(1.32)	(-2.79)	(3.96)	(1.10)	(1.54)	(-2.89)	(0.42)	(-3.56)	(-0.86)	(3.17)	(-0.18)	(0.48)	(1.44)
Adj. R^2	0.03	0.04	0.06	0.03	0.06	0.06	0.13	0.05	0.06	0.09	0.03	0.09	0.04	0.08	0.03	0.03	0.05
N	158	158	158	158	158	158	158	146	158	158	158	147	147	147	158	158	158
Panel B: VW CRSP																	
SSI_t^S	0.76*	0.72*	0.76*	0.77*	0.76**	0.95**	0.61*	0.71	0.64*	0.87**	0.75*	0.90**	0.61	0.99**	0.76*	0.76*	0.65
	(1.96)	(1.78)	(1.96)	(1.94)	(2.00)	(2.30)	(1.74)	(1.60)	(1.79)	(2.31)	(1.96)	(2.37)	(1.60)	(2.57)	(1.93)	(1.95)	(1.58)
CONT_t		0.02	0.29	0.05	0.29	-0.57*	1.25***	0.29	0.95	-0.76*	-0.09	-0.75**	-0.69	0.69**	0.22	-0.00	0.28
		(0.19)	(0.59)	(0.16)	(0.50)	(-1.87)	(3.12)	(0.65)	(1.61)	(-1.97)	(-0.16)	(-2.28)	(-1.34)	(2.00)	(0.51)	(-0.00)	(0.55)
Adj. R^2	0.03	0.02	0.02	0.02	0.02	0.03	0.10	0.03	0.06	0.05	0.02	0.05	0.04	0.04	0.02	0.02	0.02
N	158	158	158	158	158	158	158	146	158	158	158	147	147	147	158	158	158
Panel C: S&P 500																	
SSI_t^S	0.65*	0.60	0.65*	0.66*	0.65*	0.84**	0.51	0.59	0.53	0.75**	0.63*	0.77**	0.49	0.86**	0.64*	0.66*	0.56
	(1.78)	(1.59)	(1.77)	(1.75)	(1.80)	(2.13)	(1.55)	(1.41)	(1.56)	(2.10)	(1.73)	(2.14)	(1.36)	(2.35)	(1.74)	(1.78)	(1.38)
CONT_t		0.03	0.18	0.02	0.18	-0.56*	1.14***	0.29	0.96*	-0.70*	-0.20	-0.66**	-0.71	0.64*	0.28	-0.03	0.23
		(0.23)	(0.37)	(0.06)	(0.31)	(-1.87)	(2.85)	(0.68)	(1.69)	(-1.87)	(-0.39)	(-2.09)	(-1.46)	(1.91)	(0.67)	(-0.08)	(0.46)
Adj. R^2	0.02	0.01	0.01	0.01	0.01	0.03	0.09	0.02	0.05	0.04	0.01	0.04	0.04	0.03	0.02	0.01	0.01
N	158	158	158	158	158	158	158	146	158	158	158	147	147	147	158	158	158

Note: $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$

IA.7.7 Long-Short Pairs Separated

Similar to examining the long component and short component separately, it is also natural to consider each long-short ETF pair individually. Specifically, rather than calculating SSI using three leveraged-long ETFs and three leveraged-short ETFs, I analyze the return predictability arising from the S&P 500 index pair (SSO and SDS), the NASDAQ-100 index pair (QLD and QID), and the Dow Jones Industrial Average pair (DDM and DXD). I repeat the univariate regression analysis from Section IV but use each pair in place of SSI. Table IA11 provides the results.

The results in Table IA11 show that each index pair exhibits strong predictability in the univariate regressions; all three univariate regression coefficients are significant with a 5% p-value threshold (or lower) in each panel. Furthermore, the NASDAQ-100 index pair outperforms both the S&P 500 index pair and Dow Jones Industrial Average index pair in a horse race regression in which all three are included as independent variables. Albeit, in the horse race regressions, none of the coefficients are statistically significant, which is likely due to the three independent variables being strongly collinear.

IA.7.8 Institutional Ownership

My identifying assumption is that leveraged ETF share demand is relatively more sensitive to short-horizon, gambling-like demand shocks than the underlying derivative security demand. The assumption is predicated on the observation that ETF shares are traded almost exclusively by individuals and the underlying assets (i.e., derivative

Table IA11: Return Predictability and ETF Index Pairs

Regression 1 regresses the future CRSP equal weighted, CRSP value weighted, or S&P 500 index monthly returns on the net difference in each of the three ETF index pairs' share changes: $r_{t+1} = a + \beta_{SP500} SP500_t + \beta_{NASDAQ} NASDAQ_t + \beta_{DJIA} DJIA_t + \epsilon_{t+1}$ in which r_{t+1} is the future index monthly return, $SP500_t$ is the net difference in share changes from SSO and SDS, β_{SP500} is the estimated coefficient on $SP500_t$, $NASDAQ_t$ is the net difference in share changes from QLD and QID, β_{NASDAQ} is the estimated coefficient on $NASDAQ_t$, $DJIA_t$ is the net difference in share changes from DDM and DXD, β_{DJIA} is the estimated coefficient on $DJIA_t$, and ϵ_{t+1} is the error term. Regression 2 is a univariate regression using $SP500_t$: $r_{t+1} = a + \beta_{SP500} SP500_t + \epsilon_{t+1}$. Regression 3 is a univariate regression using $NASDAQ_t$: $r_{t+1} = a + \beta_{NASDAQ} NASDAQ_t + \epsilon_{t+1}$. Regression 4 is a univariate regression using $DJIA_t$: $r_{t+1} = a + \beta_{DJIA} DJIA_t + \epsilon_{t+1}$. White standard errors are used to account for heteroscedasticity and t-statistics are reported, in parenthesis, below each estimated coefficient. The sample returns run from Nov. 2006 through Dec. 2019. All variables, except for returns, are standardized.

	1	2	3	4
Panel A: EW CRSP				
SP500 PAIR	-0.59	-1.49***		
	(-1.30)	(-2.99)		
NASDAQ PAIR	-1.44		-1.71***	
	(-1.50)		(-2.91)	
DJ PAIR	0.21			-1.16**
	(0.34)			(-2.35)
Adj. R^2	0.11	0.08	0.11	0.05
N	158	158	158	158
Panel B: VW CRSP				
SP500 PAIR	-0.25	-1.06**		
	(-0.69)	(-2.43)		
NASDAQ PAIR	-1.16		-1.32**	
	(-1.23)		(-2.44)	
DJ PAIR	0.01			-0.96**
	(0.01)			(-2.45)
Adj. R^2	0.08	0.05	0.09	0.04
N	158	158	158	158
Panel C: S&P 500				
SP500 PAIR	-0.23	-0.94**		
	(-0.68)	(-2.34)		
NASDAQ PAIR	-0.98		-1.17**	
	(-1.09)		(-2.30)	
DJ PAIR	-0.04			-0.88**
	(-0.08)			(-2.42)
Adj. R^2	0.06	0.04	0.07	0.04
N	158	158	158	158

Note: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

securities) are traded by institutions. Notably, however, institutional ownership of leveraged ETF shares is not zero (see Table 2). Thus, by revealed preferences, institutions do trade leveraged ETF shares at times.

To further strengthen the case that SSI proxies for nonfundamental demand, I examine changes in institutional ownership of leveraged ETF shares in this ancillary analysis. Specifically, I utilize monthly changes in institutional ownership of leveraged ETFs using data from Bloomberg. Bloomberg reports the percentage of shares held by institutions and institutional ownership is defined as *Percentage of Shares Outstanding held by institutions. Institutions include 13Fs, US and International Mutual Funds, Schedule Ds (US Insurance Companies) and Institutional stake holdings that appear on the aggregate level. Based on holdings data collected by Bloomberg.* Similar to constructing SSI_t in equation (2), I construct net change in institutional ownership as,

$$\text{INST}_t = \sum_{i \in J} \Delta_{i,t}^{inst} - \sum_{i \in K} \Delta_{i,t}^{inst}, \tag{IA93}$$

in which J is the set of leveraged-long ETFs (QLD, SSO, DDM) and K is the set of leveraged-short ETFs (QID, SDS, DXD) and $\Delta_{i,t}^{inst}$ is,

$$\Delta_{i,t}^{inst} = \frac{\%\ \text{Ownership}_{i,t}}{\%\ \text{Ownership}_{i,t-1}} - 1. \tag{IA94}$$

Similar to SSI_t, INST_t proxies for the net demand shock for leveraged ETF shares among institutions. If the number is large and positive, institutional investors heavily demanded leveraged-long exposure via leveraged ETF shares. If the number is large and negative, institutional investors heavily demanded leveraged-short exposure.

The Bloomberg institutional ownership data is not available until early 2010. As such, I consider the ability of INST_t to predict the CRSP equal weighted index return, the CRSP value weighted index return, and the S&P 500 index return over the period May 2010 through Dec. 2019 (yielding 81 monthly observations). I perform univariate regressions of the form,

$$r_{t+1} = a\beta \text{INST}_t + \epsilon_{t+1}, \tag{IA95}$$

in which r_{t+1} is either the CRSP equal weighted index monthly return, the CRSP value weighted index monthly return, or the S&P 500 index month return in month $t + 1$, a is the regression intercept, INST_t is the net changes in institutional ownership in month t, β is the regression coefficient, and ϵ_{t+1} is the regression error term. The results are located in Panel A Table IA12.

Panel A shows that institutional demand *positively* predicts subsequent returns; a 1-standard-deviation increase in INST_t predicts a 0.84% increase in the CRSP equal weighted index the following month, a 0.74% increase in the CRSP value weighted index the following month, and a 0.67% increase in the S&P 500 index the following month. Thus, while leveraged ETF shares are rarely held by intuitions, when they are traded by institutions, those trades appear informed. This suggests that incorporating SSI may improve the measure INST. To that end, I construct,

$$\text{SSI\_MINUS\_INST}_t = \text{SSI}_t - \text{INST}_t. \tag{IA96}$$

SSI_MINUS_INST_t reflects the net demand shock among individual investors after stripping out institutional ownership changes. Panel B of Table IA12 reports the univariate return predictability regressions using SSI_MINUS_INST_t as the predictor. A 1-standard-deviation increase in SSI_MINUS_INST_t predicts a 0.92% decline in the CRSP equal weighted index the following month, a 0.79% decline in the CRSP value weighted index the following month, and a 0.72% decline in the S&P 500 index the following month.

IA.8 SSI-Managed Portfolios

The return predictability results in Section IV and the out-of-sample results in Section V.D suggest that traders could manage their portfolios based on monthly values of SSI and improve the portfolios' risk-return characteristics. In this section, I study the performance of portfolios that trade a market index based on realized values of the Speculation Sentiment Index (e.g., by using a total return swap). The three benchmark market indices considered are the CRSP equal weighted index, the CRSP value weighted index, and the S&P 500 index. I consider two basic managed portfolios: one is a long-only portfolio that only purchases a market index when SSI is negative (if SSI is positive, the portfolio does not invest in the market index and instead earns the risk-free rate) and one

Table IA12: Return Predictability with Institutional Ownership of Leveraged ETFs

In Panel A, Regressions 1-3 regresses the future CRSP equal weighted, CRSP value weighted, or S&P 500 index monthly returns on the net difference in institutional ownership for leveraged-long and leveraged-short ETFs: $r_{t+1} = a + \beta \text{INST}_t + \epsilon_{t+1}$ in which r_{t+1} is the future index monthly return, INST_t is the net difference in institutional ownership for leveraged-long and leveraged-short ETF, β is the estimated coefficient on $ints_t$, and ϵ_{t+1} is the error term. In Panel B, Regressions 1-3 regresses the future CRSP equal weighted, CRSP value weighted, or S&P 500 index monthly returns on the Speculation Sentiment Index minus the net difference in institutional ownership for leveraged-long and leveraged-short ETFs: $r_{t+1} = a + \beta \text{SSI\_MINUS\_INST}_t + \epsilon_{t+1}$ in which r_{t+1} is the future index monthly return, SSI_MINUS_INST_t is the Speculation Sentiment Index minus the net difference in institutional ownership for leveraged-long and leveraged-short ETFs, β is the estimated coefficient on SSI_MINUS_INST_t, and ϵ_{t+1} is the error term. White standard errors are used to account for heteroscedasticity and t-statistics are reported, in parenthesis, below each estimated coefficient. The sample returns run from May 2010 through Dec. 2019. All variables, except for returns, are standardized.

Panel A: INST

	1 EW CRSP	2 VW CRSP	3 S&P 500
INST_t	0.84**	0.74**	0.67**
	(2.13)	(2.10)	(2.12)
Adj. R^2	0.04	0.03	0.03
N	116	116	116

Panel B: $SSI\_MINUS\_INST_t$

	EW CRSP r_t	VW CRSP r_t	S&P 500 r_t
SSI_MINUS_INST_t	-0.92**	-0.79**	-0.72**
	(-2.39)	(-2.26)	(-2.24)
Adj. R^2	0.05	0.04	0.03
N	116	116	116

Note: * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

is a long-short portfolio that buys the market index when SSI is negative and shorts the market index when SSI is positive. Furthermore, to take advantage of large speculative demand shocks versus small ones, the notional exposures of the portfolios are determined by the magnitude of SSI.

It is worth emphasizing that the SSI-managed portfolios studied hereafter are among the *simplest* strategies one could consider.[7] A long-only SSI-managed portfolio (denoted with the superscript LO) is constructed as,

$$f_{t+1}^{\dagger,LO} = \begin{cases} -1 \times c \times \mathrm{SSI}_t \times f_{t+1} & \mathrm{SSI}_t \leq 0 \\ c \times \mathrm{SSI}_t \times rfr_{t+1} & \mathrm{SSI}_t > 0, \end{cases} \tag{IA97}$$

in which f_{t+1} is the buy-and-hold market index excess return in month $t+1$, rfr_{t+1} is the risk-free rate in month $t+1$, SSI_t is the Speculation Sentiment Index value in month t, and c is a constant that controls for the average exposure of the trading strategy. Similarly, a long-short SSI-managed portfolio (denoted with the superscript LS) is constructed as,

$$f_{t+1}^{\dagger,LS} = -1 \times c \times \mathrm{SSI}_t \times f_{t+1}. \tag{IA98}$$

For ease of interpretation and comparison, c is chosen so that each SSI-managed portfolio has the same unconditional standard deviation as its benchmark market index.[8] Note, the

[7]The motivation for simply trading based on realized values of SSI_t and not utilizing any empirical model is that it allows the entire time series (158 months) of SSI_t to be used. More sophisticated trading strategies may yield better performance but come at the cost of losing a fraction of the data in the calibration process.

[8]This is a standard assumption in the literature. See, for example, Moreira and Muir (2017).

choice of c has no effect on an SSI-managed portfolio's Sharpe Ratio. Furthermore, I do not standardize SSI to prevent any lookahead bias. As such, the value of SSI_t used in constructing the portfolio is the same as the one a trader would observe in real time.

To examine the performance of the SSI-managed portfolios, I perform time series regressions,

$$f_{t+1}^{\dagger} = \alpha + \beta f_{t+1} + \epsilon_{t+1}, \tag{IA99}$$

in which α is the intercept, $f_{t+1}^{\dagger}$ is the return on the managed portfolio, f_{t+1} is the benchmark market index excess return, β is the coefficient on f_{t+1}, and ϵ_{t+1} is the error term. A positive intercept has the economic interpretation that the SSI-managed portfolio improves the investor's Sharpe Ratio, assuming i) investors only face market risk, and ii) the benchmark market index is "the market." Table IA13 reports the results in Panel A: Regressions 1-3 report the long-only results for the whole sample, Regressions 4-6 report the long-short results for the whole sample, Regressions 7-9 report the long-only results in the post-2009 sample, and Regressions 10-12 report the long-short results in the post-2009 sample. The reported α's in Panel A are monthly values.

According to Panel A of Table IA13, in the full sample of returns, SSI-managed long-only portfolios achieve monthly α between 0.60%-0.71% (7.20%-8.47% annually) and the long-short portfolios achieve monthly α between 0.80%-0.97% (9.57%- 11.66% annually). Panel A also provides the annualized appraisal ratios (i.e., excess return Sharpe

Table IA13: Trading Strategy Performance

In Panel A, a SSI-managed portfolio's return is regressed on the contemporaneous monthly return of either the CRSP equal weighted index, CRSP value weighted index, or S&P 500 index monthly index: $f^{\dagger}_{t+1} = \alpha + \beta f_{t+1} + \epsilon_{t+1}$ in which $f^{\dagger}_{t+1}$ is the SSI-managed portfolio return, α is the SSI-managed portfolio's alpha, f_{t+1} is the reference index monthly excess return, β is the estimated coefficient on f_{t+1}, and ϵ_{t+1} is the error term. In Regressions 1-3 and 7-9, SSI-managed portfolios are long-only strategies. In Regressions 4-6 and 10-12, SSI-managed portfolios are long-short strategies. Panel A also reports the RMSE from the regressions and the SSI-managed portfolio's annualized appraisal ratio ($\frac{\alpha\sqrt{12}}{\sigma_\epsilon}$). In Panel B, SSI-managed portfolio alphas are reported, controlling for the Fama–French 3-factors (Mkt, SMB, and HML). In Panel C, realized annual Sharpe Ratios are reported for each SSI-managed portfolio and the reference index Sharpe Ratios are included for reference. In Regressions 1-6, the sample returns run from Nov. 2006 through Dec. 2019. In Regressions 7-12, the sample returns run from Jan. 2010 through Dec. 2019. White standard errors are used to account for heteroscedasticity and t-statistics are reported, in parenthesis, below each estimated coefficient.

Panel A: Univariate Regressions

	Full Sample						Post-2009					
	Long Only			Long-Short			Long Only			Long-Short		
	1	2	3	4	5	6	7	8	9	10	11	12
	EW†	VW†	SP500†	EW†	VW†	SP500†	EW†	VW†	SP500†	EW†	VW†	SP500†
Alpha (α)	0.71**	0.63**	0.60**	0.97**	0.88*	0.80*	0.81***	0.59**	0.51**	0.96**	0.77**	0.66*
	(2.33)	(2.22)	(2.11)	(1.99)	(1.82)	(1.80)	(3.15)	(2.61)	(2.22)	(2.40)	(2.12)	(1.97)
EW	0.53***			-0.19			0.59***			0.09		
	(3.61)			(-0.74)			(4.64)			(0.52)		
VW		0.51***			-0.23			0.59***			0.12	
		(4.18)			(-0.86)			(4.28)			(0.68)	
SP500			0.53***			-0.21			0.60***			0.16
			(4.51)			(-0.83)			(4.43)			(0.96)
Adj. R^2	0.28	0.26	0.28	0.03	0.05	0.04	0.34	0.34	0.35	0.00	0.01	0.02
N	158	158	158	158	158	158	120	120	120	120	120	120
RMSE	4.19	3.70	3.53	4.85	4.20	4.08	3.22	2.96	2.88	3.97	3.63	3.54
APPRAISAL	0.58	0.59	0.59	0.69	0.72	0.68	0.87	0.69	0.61	0.84	0.73	0.65

Panel B: Alphas Controlling for Fama–French 3-Factors

	1	2	3	4	5	6	7	8	9	10	11	12
Alpha (α)	0.90**	0.84**	0.80**	1.01*	0.74*	1.02**	0.79***	0.56**	0.97***	0.92**	0.63**	1.16**
	(2.46)	(2.16)	(2.47)	(1.96)	(1.87)	(2.02)	(2.66)	(2.22)	(2.76)	(2.29)	(2.00)	(2.50)

Panel C: Realized Sharpe Ratios

	1	2	3	4	5	6	7	8	9	10	11	12
Strategy SR	0.70	0.79	0.74	0.60	0.58	0.56	1.09	1.11	1.01	0.89	0.83	0.78
Index SR	0.44	0.61	0.52	0.44	0.61	0.52	0.69	0.97	0.92	0.69	0.97	0.92

Note: $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$

Ratios) for each of the portfolios. The annualized appraisal ratio is calculated as,

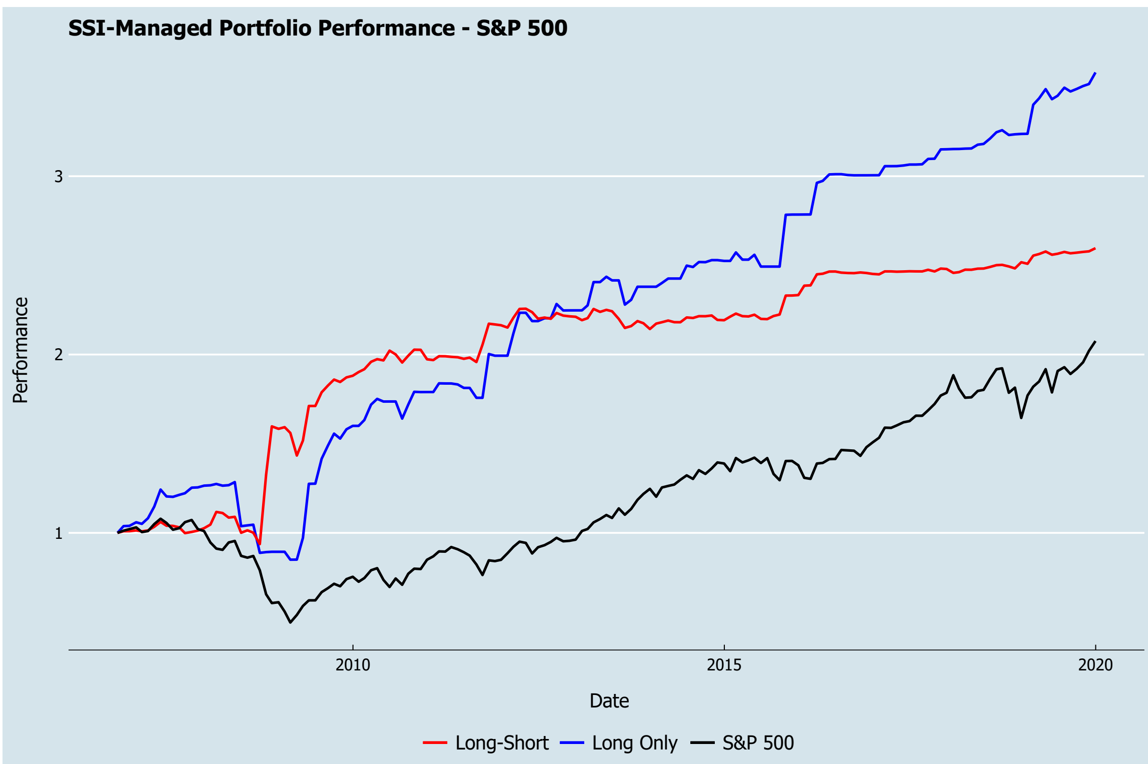
$$\text{Annualized Appraisal Ratio} = \frac{\alpha\sqrt{12}}{RMSE}, \tag{IA100}$$

and has the interpretation that it is the portfolio's excess return per unit of risk. For the long-only portfolios, the appraisal ratios are between 0.58-0.59 for each benchmark market index. For the long-short portfolios, the appraisal range between 0.68-0.72. The post-2009 sample yields monthly portfolio α's that are qualitatively similar and statistically significant. Furthermore, the annualized appraisal ratios are in the same range as the full sample results.

Panel B of Table IA13 provides estimates of α after controlling for the Fama–French 3-factors (Fama and French (1993)). The α estimates are qualitatively the same in economic magnitude and statistical significance. Panel C of Table IA13 dovetails with the reported values of α and reports the annualized Sharpe Ratios for each of considered portfolios and compares it to the portfolio's benchmark index Sharpe Ratio over the same period; SSI-managed portfolios typically achieve superior Sharpe Ratios. Finally, Figure IA1 depicts the performance of one dollar invested in one of three portfolios: the S&P 500 Index, a long-only SSI-managed S&P 500 portfolio, and a long-short SSI-managed S&P 500 portfolio. As can be see in the graph, SSI-managed portfolios outperform the benchmark index. Figures showing the performance of SSI-managed CRSP equal weighted index portfolios and SSI-managed CRSP value weighted index portfolios look qualitatively the same.

Figure IA1: Trading Strategy Performance

Figure IA1 shows the performance of one dollar invested in one of three portfolios: the S&P 500 Index, a long-only SSI-managed S&P 500 portfolio, and a long-short SSI-managed S&P 500 portfolio. The sample covers Nov. 2006 through Dec. 2019.



IA.9 Correlation between SSI and Control Variables

In this ancillary analysis, I provide the correlations between SSI and the control variables used in the empirical analysis. Specifically, I calculate pairwise correlations between SSI and index monthly returns (r_{ew}, r_{vw}, and r_{sp}), cyclically adjusted earnings-to-price (CAEP), term spread (TERM), dividend-to-price (DP), short-rate (RATE), variance risk premium (VRP), intermediary capital risk factor (INTC), innovation to aggregate liquidity (Δ_LIQ), short interest (SHORT), VIX (VIX), Baker-Wurgler sentiment level (SENT), aligned investor sentiment level (HJTZ), closed-end fund discount (CEFD), consumer confidence level (CONF), change in consumer confidence (Δ_CONF), and investor lottery demand (FMAX). The results are located in Table IA14.

The results in Table IA14 provide additional evidence that SSI is contrarian in nature: SSI is strongly negatively correlated with contemporaneous returns (r), the variance risk premium (VRP), the intermediary capital risk factor (INTC), innovations to aggregate liquidity (Δ_LIQ). Indeed, each of these control variables is associated with poorly performing markets. To that end, SSI is strongly positively correlated with monthly VIX (VIX). Thus, when VIX levels are high (which typically occur in down markets), SSI is bullish.

Table IA14: Correlations of SSI with Control Variables

Table IA14 presents the correlation coefficients for SSI and the controls used throughout the analysis: index monthly returns (r_{ew}, r_{vw}, and r_{sp}), cyclically adjusted earnings-to-price (CAEP), term spread (TERM), dividend-to-price (DP), short-rate (RATE), variance risk premium (VRP), intermediary capital risk factor (INTC), innovation to aggregate liquidity (Δ_LIQ), short interest (SHORT), VIX (VIX), Baker-Wurgler sentiment level (SENT), aligned investor sentiment level (HJTZ), closed-end fund discount (CEFD), consumer confidence level (CONF), change in consumer confidence (Δ_CONF), and investor lottery demand (FMAX). The sample returns run from Nov. 2006 through Dec. 2019 (if the control variable is available through 2019). The t-statistics are calculated as $t = r\sqrt{\frac{n-2}{1-r^2}}$, in which r is the sample correlation and n is the number of paired observations. Statistical significance is determined using a Student's t-distribution with degrees of freedom of $n - 2$.

Correlations with SSI

	SSI
r_{ew}	-0.66***
r_{vw}	-0.63***
r_{sp}	-0.60***
CAEP	0.01
TERM	0.04
DP	0.05
RATE	-0.02
VRP	-0.45***
INTC	-0.49***
Δ_LIQ	-0.27***
SHORT	0.13
VIX	0.31***
SENT	0.09
HJTZ	0.30***
CEFD	0.07
CONF	-0.11
Δ_CONF	-0.08
FMAX	-0.47***

Note:

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

References

Brown, D. C.; S. W. Davies; and M. Ringgenberg. "ETF arbitrage, non-fundamental demand, and return predictability." *Review of Finance*, (2021).

Evans, R. B.; R. Moussawi; M. S. Pagano; and J. Sedunov. "ETF failures-to-deliver: Naked short-selling or operational shorting?" *Working Paper*, (2019).

Fama, E. F., and K. R. French. "Common risk factors in the returns on stocks and bonds." *Journal of Financial Economics*, 33 (1993), 3–56.

Ivanov, I. T., and S. L. Lenkey. "Are concerns about leveraged ETFs overblown?" *Federal Reserve Board: Finance and Economics Discussion Series*, (2014).

Moreira, A., and T. Muir. "Volatility-managed portfolios." *The Journal of Finance*, 72 (2017), 1611–1644.

Staer, A. "Fund flows and underlying returns: The case of ETFs." *International Journal of Business*, 22 (2017).

Stambaugh, R. F. "Predictive regressions." *Journal of Financial Economics*, 54 (1999), 375–421.